Application of Multivariate Geospatial Statistics to Soil Hydraulic Properties

---

By: James Lea Bevington

Under direction of George Vellidis and Francesco Morari

Abstract

Currently, an inability to accurately measure and model hydraulic properties of soil at large scales plagues the accuracy of climate and crop modeling. Much effort is focused on development of new models and pedotransfer functions which include new variables and soil processes. However, soil is inherently spatially dependent. Therefore, many soil properties and relationships can be analyzed with geospatial statistics. In this thesis, a multivariate geospatial technique called factorial kriging analysis (FKA) was employed to analyze the spatial relationships of soil hydraulic parameters. FKA was also used to develop a novel prediction method to aid in characterizing the spatial distribution of an area's hydraulic properties. Bulk density was found to be the source of short scale variance while paleo-channels were found to be the major source of variance at long scale. The prediction method described within is able to classify hydraulic zones and distinguish between groups of water retention curves.

Index Words: Factorial Kriging Analysis, Multivariate geospatial statistics, Unsaturated hydraulic conductivity, soil hydraulic properties, soil water retention, van Genuchten model, Mualem model, Paleo-channel, Peat, Venice Italy, Saltwater intrusion

APPLICATION OF MULTIVARIATE GEOSPATIAL STATISTICS TO SOIL HYDRAULIC PROPERTIES

By

James Lea Bevington

B.S., The University of Tennessee, 2011

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirement for the Degree

MASTERS OF SCIENCE

ATHENS, GEORGIA

2013

APPLICATION OF MULTIVARIATE GEOSPATIAL STATISTICS TO SOIL HYDRAULIC PROPERTIES

By

James Lea Bevington

Major Professor: George Vellidis

Committee: Francesco Morari
Christof Meile
David Radcliffe

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2013

Acknowledgments

When I started my masters at UGA I couldn't have imagined all that I would learn and accomplish in the coming years. I owe many thanks to those who have been part of the journey. First and foremost I want to thank all of the folks at TAPAC for developing the exchange program which funded my assistantship and allowed me to experience international life. Funding for my assistantship came from the USDA NIFA International Science and Education Program to the University of Georgia (Award No. 2010-51160-21050) to support TAPAC graduate student exchanges. Funding for the overall research project was provided by GEO-RISKS: Geological, Morphological and Hydrological Processes: Monitoring, Modeling, and Impact in the North-Eastern Italy, WP4, University of Padova, Italy. Without these funding contributions, this thesis would not be possible.

The lessons I learned while living abroad and interacting with an international group go well beyond the classroom and any of the technical content of this thesis. In particular, I am grateful to Dr. Vellidis for taking a risk on me and trusting my capability. I also appreciate Dr. Morari for envisioning and outlining this research, providing his input and speaking English with me even when he was tiered. I am thankful to Dr. Meile for listening to my wild ideas as a potential graduate student and arranging the meetings that ultimately lead to my enrollment at UGA. To Dr. Radcliffe I show gratitude for helping me learn the soil side of my research before leaving to go to Italy. As a whole, my committee deserves recognition for collaborating under difficult, remote circumstances and for their effort in keeping me focused and on track. I owe thanks to Elia for all of his help related to the project.

Ringrazio Chiara per tutto l'assistenza che lei ha dato con amicizia, le spiegazioni della vita italiana, e l'aiuto con i documenti. Senza lei, sarei stato esiliato da Italia. Grazie anche ai miei amici che

mangiano con me. Senza loro, avrei mangiato da solo ogni giorno. In più Giovanni chi ha dimostrato Padova per me su mi primi giorno. Italia ha una bella cultura e sono grato che abitavo là.

Translation: Beyond the technical parts of my work, I am grateful to Chiara for keeping me company, explaining Italian life and going through the visa process with me. I probably would have been deported without her help. I am thankful to the “lunch crew” for making me a part of their group. Without them, I would have to eat lunch alone. Giovanni, who showed me around Padova my first day, also deserves mention. Italy has a beautiful culture and I am thankful to have lived there.

I apologize to the reader for the ridiculous format of this thesis. This format is required by the graduate school and apparently they like it. It makes my eyes hurt but sometimes you concede to bureaucracy to get you degree.

Table of Contents

Chapter 1: Introduction and Literature Review

1. Motivation for this Study

Climate change is one of the most important challenges facing our global society and its impacts are diverse and far-reaching. This study will address two components of climate change: the effects of sea-level rise on coastal agriculture and large scale modeling of the vadose zone. Due to sea-level rise, coastal crop land is being threatened by a phenomenon called saltwater intrusion which is the process of denser saltwater replacing freshwater below the soil surface. As sea-level rises, the inland distance that saltwater can intrude increases because of changes in the hydraulic gradient and previously unaffected land develops salinity problems. Loss of productivity of agricultural land has implications for the local and global food supply.

The vadose zone is an important component of the soil profile and serves as the interface between groundwater, terrestrial vegetation and the atmosphere. Consequently, inclusion of vadose zone processes in climate modeling provides an opportunity for improved accuracy of climate change predictions. A limiting factor in including vadose zone processes in climate models is an inability to collect enough data points by direct methods and inadequate prediction models.

One important challenge in vadose zone research is the discrepancy between lab measured and field measured unsaturated hydraulic properties. The problem arises from an inability to define accurate boundary flux conditions for field sites which leads to difficulties in modeling and poor declaration of error sources.

2. Saltwater Intrusion

Saltwater intrusion is the process of saltwater displacing freshwater in the water table. Figure 1.1 illustrates the basic principal. This is a direct result of the higher density of saltwater compared to freshwater density. The higher density creates a higher static pressure at the same column height than the freshwater. Because the saltwater is seaward and the freshwater landward, the result is movement of saltwater in a landward direction. The distance that the saltwater moves inward is controlled by the difference in column height and density between the saltwater and freshwater. Typically, the elevation of the water table increases with inland distance from the sea which adds to the column height of the freshwater compared to the saltwater which remains at sea-level. However, an increase in sea-level adds to the column height of the saltwater and results in a landward movement of saltwater. Saltwater intrusion also occurs in river mouths and is mostly governed by the same principals. The leading exception is the case of mixing where the interface is not sharp. This usually occurs in tidal systems (Custodio and Bruggeman, 1987).

Other factors influencing saltwater intrusion include land subsidence and removal of fresh water from ground water aquifers. As land subsides and surface elevation decreases, often the water table elevation also decreases. This results in increased saltwater intrusion. Use of fresh groundwater for irrigation, industrial purposes and other uses reduces the column height of the freshwater. This can lead to saltwater contamination of the aquifer and loss of the freshwater source.

The Ghyben-Herzberg equation can be used to estimate saltwater intrusion though it is not accurate in all cases. It is given by $z_s = \frac{\rho_f}{(\rho_s - \rho_f)} z_f$ where z_s is the depth below sea-level of the saltwater-freshwater interface, z_f is the height above sea-level of the water table surface, and ρ_s and ρ_f are saltwater and freshwater densities (Custodio and Bruggeman, 1987).

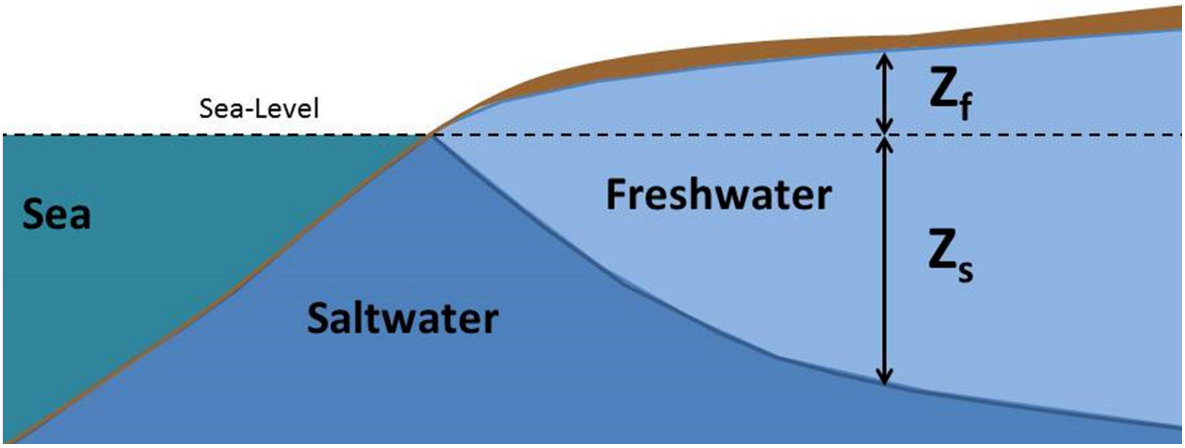


Figure 1.1. Illustration of saltwater intrusion. Z_f is the height of the freshwater surface above sea-level, Z_s is the distance from lea-level to the saltwater surface. As sea-level rises, Z_s for any given distance inland decreases bringing saltwater closer to the surface.

3. Unsaturated Hydraulics of Soil

3.1 Water Retention in Soil

Water properties of soil are dominated by adhesion and cohesion, both of which are caused by the polar shape of the water molecule. Adhesion and cohesion act together resulting in capillarity which is the pull of water as a result of being in a narrow tube. The magnitude of this force is a function of the radius of the tube where an increase in radius decreases the capillary force. In soil, this radius is not constant or even measurable, but the concept is useful in understanding its hydraulic characteristics. The basic idea is that water molecules in a small pore will be held with a stronger force than those in a large pore. Therefore, large pores will empty before smaller pores assuming there is not a restriction on flow.

The most widely used soil water retention curves are van Genuchten (1980) and Brooks and Corey (1964). There are others which are best applied in unique situations such as Ross et al. (1991) for extremely dry conditions. The Brooks and Corey model has two major draw backs. There is a sharp

discontinuity near saturation and no inflection point. One reason that it is valuable to have smooth functions is so that it can be differentiable. Taking the derivative of the water retention curve yields a plot which is easily converted to a frequency distribution of pore size (Radcliffe and Simunek, 2010). This function is called the hydraulic capacity function. The van Genuchten model has an inflection point and improves on saturation performance. The disadvantages of the van Genuchten include its potentially poor fit at high potentials (Assouline and Tartakovsky, 2001) and the limiting constraints which must be used for compatibility with many of the conductivity functions. Many other models with potentially better performance have been developed but most have not caught on (Radcliffe and Simunek, 2010). One model of particular note is the Dual-Porosity model of Durner (1994) which extends the van Genuchten model to fit bimodal pore distributions. While the Durner model offers improved prediction accuracy, it introduces several more parameters which may be limiting in their own way. The van Genuchten model is given by $\theta(h) = \frac{\theta_s - \theta_r}{[1+(-\alpha h)^n]^m} + \theta_r$ where θ_s is the water content at saturation, θ_r is the residual water content, α is related to the air entry point and *n* and *m* are fitting parameters. Typically the constraint of $m = 1 - \frac{1}{n}$ is added in order to obtain a closed form expression (van Genuchten, 1980). A plot of this model can be seen in figure 1.2A. A useful modification of the van Genuchten model is to normalize the curve by removing $\theta_s \; and \; \theta_r$. This version is called the effective saturation curve and is given as $Se(h) = \frac{1}{[1+(-\alpha h)^n]^m}$ and is depicted in figure 1.2B.

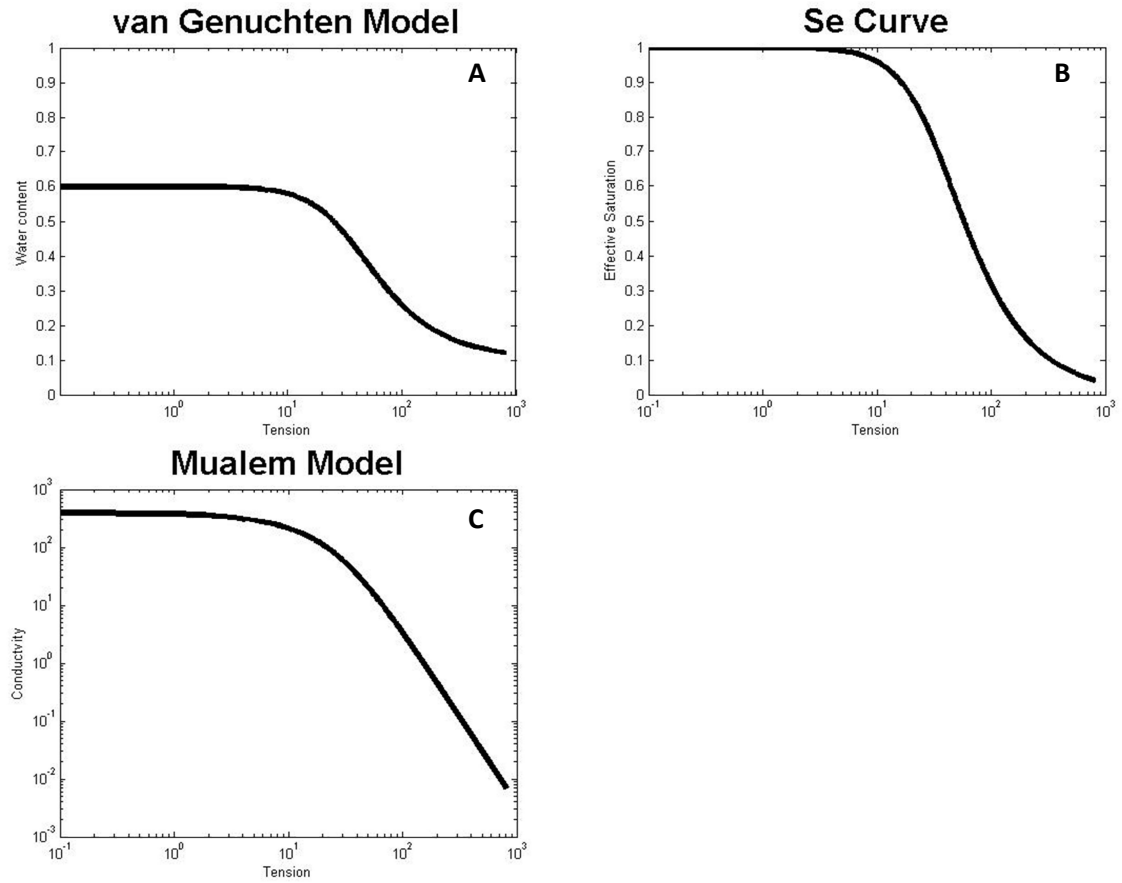


Figure 1.2A-C. A) van Genuchten model for WRC. The Y intercept is the water content at saturation. As tension increases (moving to the right) water is removed resulting in a negative slope B) Effective saturation curve. This is a normalized WRC and the input for the Mualem UHC model. C) Mualem UHC model. The Y intercept is K_s. Increasing tension removes water which reduced cross-sectional area resulting in a decrease in conductivity. Notice the similarity in curve shape near saturation.

3.2 Water Movement and Unsaturated Hydraulic Conductivity

A commonly used equation to model water movement through soil in the saturated state is Richards' (1931) equation given by $\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z} + 1\right)\right]$ where $K(\theta)$ is unsaturated hydraulic conductivity, θ is water content, z is elevation above a datum, t is time, and h pressure head. The derivation of this equation starts with Darcy's law for saturated flow through porous medium (Richards, 1931). As water is removed, the cross sectional area of the water portion becomes smaller and the force required to remove water becomes larger. These changes are accounted for in the conductivity function

which relates conductivity to water content. Substituting a water retention function, conductivity can be expressed as a function of tension. Richards' equation is a nonlinear partial differential equation which has no analytical solution. Therefore most applications of Richards' equation involve numerical approximation (Radcliffe and Simunek, 2010).

There are two popular conductivity models, Burdine (1953) and Mualem (1976). Typically, the Mualem model is used in conjunction with the van Genuchten retention model while Burdine's is used with Brooks and Corey. In a study by van Genuchten and Nielsen, (1985) the van Genuchten-Mualem combination gave the best results for a wide range of soils. The Mualem conductivity curve in terms of van Genuchten parameters is $K(Se) = K_s Se^l[1 - \left(1 - Se^{1/m}\right)^m]^2$ where $m = 1 - \frac{1}{n}$ and $n > 1$ and l is the pore connectivity parameter and K_s is the conductivity at saturation. An example of this model is given in figure 1.2C.

3.3 Application of Models to Data

As with many PDEs there are two methods to fitting models to measured data, forward and inverse. A forward method consist of a nonlinear regression while an inverse method uses data as boundary conditions then iteratively solves for the parameters that fit the data. For unsaturated hydraulic studies, $K(h)$ is usually not measured directly. Most methods measure the change in mass over time and the tension at multiple locations in the sample over time to establish a pressure gradient. In order to obtain $K(h)$ points for the forward method, the gradient is assumed to be linear which simplifies the system and allows for a direct calculation. For the most part, this assumption tends to be valid for weak gradients but becomes stretched as the gradient increases. Using the inverse method avoids this assumption altogether because it does not require input of $K(h)$ points.

Solving Richards' equation numerically can be accomplished in a plethora of ways including finite difference, finite element, and finite volume (van Genuchten, 1982; Manzini and Ferraris, 2004).

The most commonly used code is HYDRUS 1D (Simunek et al., 2008) which is a mass-lumped linear finite elements method. It uses the water flux data and given hydraulic function parameters to predict the tension at the nodes over time. Then it compares the predicted tensions to the measured values. The accuracy is evaluated with an objective function that is minimized using the Marquardt-Levenberg scheme. This scheme is a local method which means that the solution may not be the global minimum; accuracy is best when the minimum is well defined. The objective function was evaluated by Simunek et al. (1998) for its performance and the effect of parameter error on the shape of the function. They found that failure to minimize the objective function was mostly related to inaccurate initial guesses in θ_r and sometimes n while minima were best defined for θ_s and n. θ_r had a broadly defined minimum. All parameters except α improved with increased experimental time.

4. Spatial Statistics

4.1 Variography

In geospatial statistics the value of a property is taken as the result of an underlying random process which is considered to be the source of spatial variability. For soil, this framework fits well because soil is largely explained by five formation processes (Brady and Weil, 2008; Goovaerts, 1992; Goovaerts, 1999). The theory of regionalized variables was pioneered by Georges Matheron beginning in 1962 (Chiles and Delfiner, 2012). The basic idea is given as Tobler's first law (1970) of geography, "Everything is related to everything else, but near things are more related than distant things" or in statistical terms, a variable is said to be autocorrelated which means that it is correlated with itself. Autocorrelation is described by the variogram $\gamma(h) = E[Z(x+h) - Z(x)]^2$ where h is the separation distance between two points, Z(x) is a stationary random function and x is point in space. There are two assumptions on Z(x); first, its mean is stationary in space i.e. no global trends, and second its variance is stationary in space. If global trends are present they can be removed by fitting a polynomial trend to the

data then using the residuals for further analysis. Typically, an experimental variogram is derived from measured point data. The two point statistic is calculated for each pair of points in the collected data then binned into predefined h-lags. The values for each lag are determined by averaging the points for each bin. A variogram model can then be fit to the experimental variogram. There are several variogram models including, linear, exponential and stable. However, the spherical model is the most common and the best for illustrating the concepts; an example is seen in figure 1.3A. The spherical model is given by $\gamma(h) = a + (s - a) - \frac{3}{2}\frac{h}{r} + \frac{1}{2}\frac{h^3}{r^3}$ if $0 \leq h \leq r$ and $\gamma(h) = \sigma^2$ if $h \geq r$ where *a* is the nugget, *s* is the sill, and *r* is the range. The nugget is the y intercept and its presence indicates poor spatial structure either from low autocorrelation or high degree of error in measurements. The sill is the maximum y value and should approximate the sample variance. The separation distance where the model meets its sill is the range and indicates the maximum distance that values are autocorrelated. Often times, there are more than one active processes which require a nested variogram model like the one in figure 1.3B $\gamma(h) = \gamma_1(h) + \gamma_2(h) + \cdots$. For multivariate analysis, a cross-variogram can be calculated using a similar methodology to the univariate case where the cross-variogram is $\gamma_{ij}(h) = E\big[(Z_i(x+h) - Z_i(x)\big) \cdot (Z_j(x+h) - Z_j(x))]$. One added constraint to a multivariate study is that given by the linear model of coregionalization (LMC) which requires that the same set of ranges be used in all variogram and cross-variogram models (Wackernagel, 2003). This requires a simultaneous fitting of all models but ensures that the resulting model will be positive definite- a necessary constraint for subsequent analysis. The resulting model is informative in its self because it illustrates the spatial covariance between variables. Yet, as in descriptive statistical analysis, there is much more information to be explored.

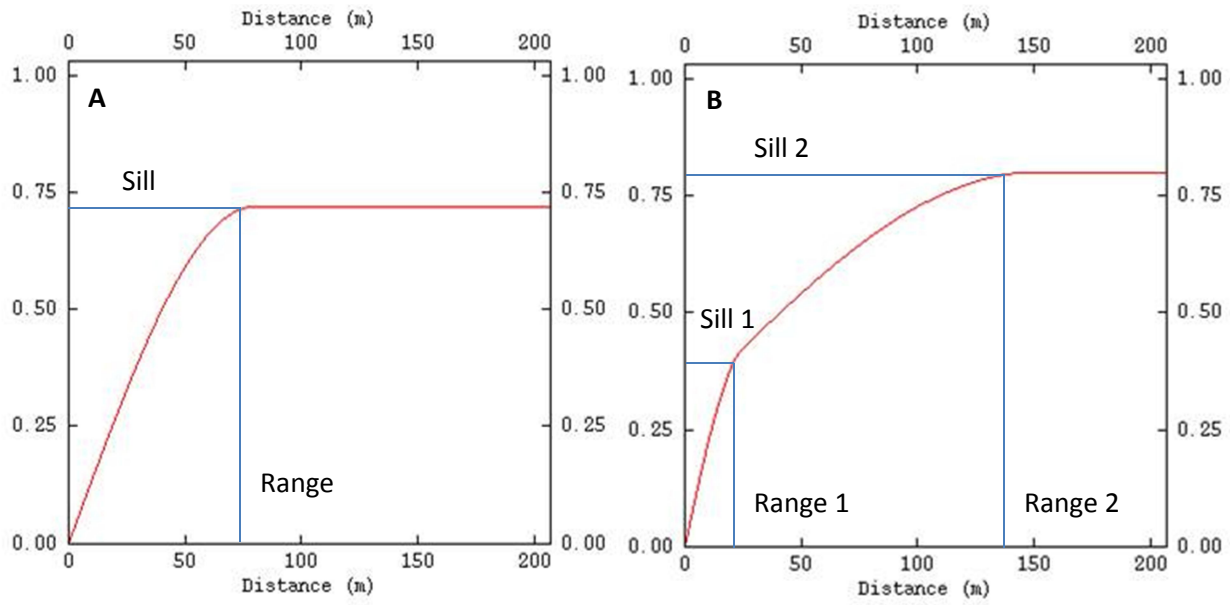


Figure 1.3A-B. A) Spherical variogram model. The X axis represents the distance separating two points and the Y axis represents variability in attributes of these points. The sill is the maximum variability found in the phenomena under study and the range is the separation distance where the sill is reached. B) Two nested spherical models. The nested model is simply two models added together which is common in soil science. Each has a distinguished range and sill.

4.2 Kriging

One application of spatial statistics is the construction of continuous variable maps from point measured values. This is accomplished by interpolating values at unvisited locations from values at known locations. In kriging interpolation, the variogram model is used to assign weights to known values based on their distance from the unvisited location. The weights are used in a weighted averaging scheme where nearby points are assigned higher weights than farther away points. For each unvisited location, new weights are assigned and the process is repeated. One subjective decision that the user must make is that of neighborhood size which determines the points that will be included in the averaging for an unvisited site. A large neighborhood will include more points but the points farther away will be less related and may result in an overly smooth surface. The kriging system is given by $Z^* = \sum_1^s w_s \cdot Z_s$ where Z^* is the estimated value at the unvisited point, and w_s is the assigned weight for

the known value Z_s, at point *s*. For the multivariate case, the cokriging system is given by $Z^* = \sum_1^v \sum_1^{s_v} w_s^n \cdot Z_s^v$ where *v* is the number of variables. The neighborhood for the cokriging system is also more complicated because the number of samples for each variable may be different (Wackernagel, 2003). If each variable is defined at all sample points then the neighborhood is called isotropic; otherwise, it is heterotopic. The ability to have a differing number of samples is empowering because it enables the inclusion of secondary variables. A secondary variable is one that relates to the primary variable but is easier to measure thereby enabling the inclusion of more samples to inform interpolation. In many soil related studies, remotely sensed data which are known for the entire study site are utilized as the secondary variable. In the case of exhaustively known secondary variables, the use of a special type of cokriging called multicollocated cokriging can greatly reduce computational efforts. In multicollocated cokriging, the neighborhood is almost isotropic except that the secondary variable remains defined at the unvisited location being estimated. For each unvisited location the neighborhood changes to include the secondary variable at that location. While the number of data points used in a single realization is drastically reduced, the amount of information lost is minimal due to the shielding effect of the highly weighted secondary point at the unvisited location (Rivoirard, 2001).

4.3 Multivariate Geostatistics

Using the variogram model, further structural information can be extracted with additional analyses in the multivariate case. For one, regionalized correlation coefficients or coefficients of codispersion can be calculated by $R_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(h) \cdot \gamma_{jj}(h)}}$ where $R_{ij}(h)$ is the coefficient and γ is the respective variogram value at *h*. Setting *h* equal to a range value yields an indication of which variables are related through the process that acts at that scale. Regionalized correlation coefficients are often much stronger than global correlation coefficients. An example of this is given in Goovaerts (1998a) where the short scale coefficient between two variables was -0.31 and the long scale was 0.99 yet the

global coefficient was much less informative because the opposing processes cancel. Even beyond this, a regionalized PCA can be conducted at each scale to produce representative PCs which describe the majority of the variance for the scale. Interpolating these PCs produces maps that depict the spatial distribution common amongst highly loading variables to the PC. Visual inspection of these maps often leads to identification of the underlying processes. This form of a regionalized PCA is better known as Factorial Kriging Analysis (FKA) given by $Z_k^* = \sum_{u=1}^{v} a_{ku} \cdot Y_k^u$ where *k* is the index representing the spatial structure, *u* is the index representing the regionalized PC, Y_k^u is the set of PCs and a_{ku} are weighting coefficients which are the elements of matrix $\mathbf{A}_k$ which is derived by PCA of the regionalized covariance matrix (Chiles and Delfiner, 2012).

5. Secondary Variables

5.1 Apparent Electrical Conductivity

Remotely sensed apparent electrical conductivity (ECa) has become a popular tool in soils and precision agricultural studies. One reason is the plethora of information held in a single ECa measure, while other reasons include its ease of use, inexpensive application, and the ability to rapidly collect thousands of data points. These qualities make ECa an obvious choice for geospatial studies and have the potential to aid in large scale hydrologic applications (Robinson et al., 2008). The electrical conductivity of soil is made up of the conductivity of three sub-paths, the liquid water in the soil, the solid particles which are in direct contact and a combination of both (Corwin and Lesch, 2005). ECa is related to these through $ECa = \left[\frac{(\theta_{ss}+\theta_{ws})^2 EC_{ws} EC_{ss}}{\theta_{ss} EC_{ws} + \theta_{sw} EC_s}\right] + (\theta_w - \theta_{ws}) EC_{wc}$ where θ is water volumetric content and EC is the specific electrical conductivity (Rhoades et al. 1989). Some simplifications include $\theta_w = \frac{P_w \rho_b}{100}$, $\theta_{ws} = 0.639\theta_w + 0.011$, $\theta_{ss} = \frac{\rho_b}{2.65}$, $EC_{ss} = 0.019(s_p) - 0.434$, and $EC_W = \frac{EC_e \rho_b s_p}{100\theta_w}$ where P_w is mass percent of water, ρ_b is bulk density, s_p is saturation percentage, and ECe is the electrical

conductivity of the soil (Corwin and Lesch, 2005). From a macroscopic point of view, ECa is affected by soil salinity, soil water content, effective saturation, and bulk density. These properties may also be functions of other properties such as texture, organic content, and soil structure. Even though all of these variables are influential, typically a study site will be dominated by 1-2 of variables which is important to consider when using ECa (Corwin and Lesch, 2005).

5.2 Peat Oxidation and Subsidence

The reclaimed region of the study site is known to have peat soils that are rich in organic content. Prior to reclaiming and draining the land, this organic content was more or less protected because it was underwater. Now that it is exposed to the atmosphere, oxidation converts the organic matter into CO_2 which consolidates the material and results in land surface subsidence. The rate of oxidation is influenced by water content and land use practices-namely agriculture (Gambolati et al., 2006). Gambolati et al. (2006) studied these phenomena in the region and verified that peat oxidation and elevation are strongly correlated. They confirm the claims by Carbognin et al. (2005) of 2-3cm/yr subsidence rates.

6. Project Description

6.1 Description of Overall Project

The overall project goal is to develop and validate a procedure based on experimentation and modeling to better understand saltwater intrusion and its effects on agriculture in the Po River plain. This knowledge will inform management of the area for improved agricultural production. Specific objectives include: 1. defining the physical and chemical geologic structure of the farmland, 2. quantify the temporal and spatial distribution of saltwater intrusion, 3. calibrate and validate flow and transport models, 4. integrate remote sensing and precision agricultural techniques with crop stress from

saltwater and 5. develop and validate a model to simulate local scale crop yield from water movement in the unsaturated zone. Much of this work is useful not only to the Po River plain but also to many other settings because several locations around the world are experiencing saltwater intrusion.

As this is a multi-disciplinary project, collaborators from several research groups and institutions are contributing. The involved institutions include the Department of Mathematical Methods and Models for Scientific Applications, the Department of Environmental Agronomy, and the Department of Geosciences at the University of Padova, Italy, and the Institute of Marine Sciences, National Research Council, Venice, Italy.

This study offers many opportunities not only to investigate high impact global problems but also to address some of the most pressing issues in the field of soil science. This research involves the prediction of crop yield from field scale water movement via a coupled saturated and unsaturated model. The vadose zone is an intermediary between the saturated boundary and the crops so accuracy of yield predictions depends on accuracy the vadose zone model. The other intermediary is the prediction of yield based on salinity, but this prediction is not spatially dependent and its error can be well defined. Therefore, mapping the model error will likely reveal locations of accurate and inaccurate prediction based mostly on errors in the vadose zone component of the model. From there, these errors can be further studied and understood.

6.2 Description of the Current Work

The study site used for this work is a 26.5 ha field located on the southern shores of the Venetian Lagoon (figure 1.4). Due to wet conditions, a portion of the study site was abandoned leaving only 20.8 ha. This area was reclaimed from the lagoon by the construction of levees between 1892 and 1967 (Gambolati et al., 2006). A pumping station and a dense network of ditches control the depth to the water table, which is generally maintained at ca. 0.6 m during the summer season in order to

promote sub-irrigation. On the northern edge are the Morto Canal, Bacchiglione River, Brenta River and Venetian Lagoon. All of the canals and rivers are not natural as the rivers were diverted here at various times in the past. During construction of the Morto Canal, soil was placed on the northern edge of the field burying the original surface Because of its reclaimed origin, soil shows a high variability in terms of texture, pH, salinity and presence of organic horizons, which can be superficial as well as buried at deeper depth. Two main soil typologies can be identified according to USDA (1998), in the northern part, along the Morto Canal, Fluvaquentic Endoaquepts fine-silty, mixed, calcareous, mesic; in the southern part Typic Sulfisaprists euic, mesic. Moreover, there are well preserved sandy paleo-river channels which cut through the field. These are artifacts of its geological history as a river delta. In recent years, the field has been continuously used to grow corn (*Zea mais* L.) and is tilled to a depth of approximately 30 cm each year. One advantage of this site is the wide variability of the soil. This variability introduces a wider range of soils and therefore wider validity of results.

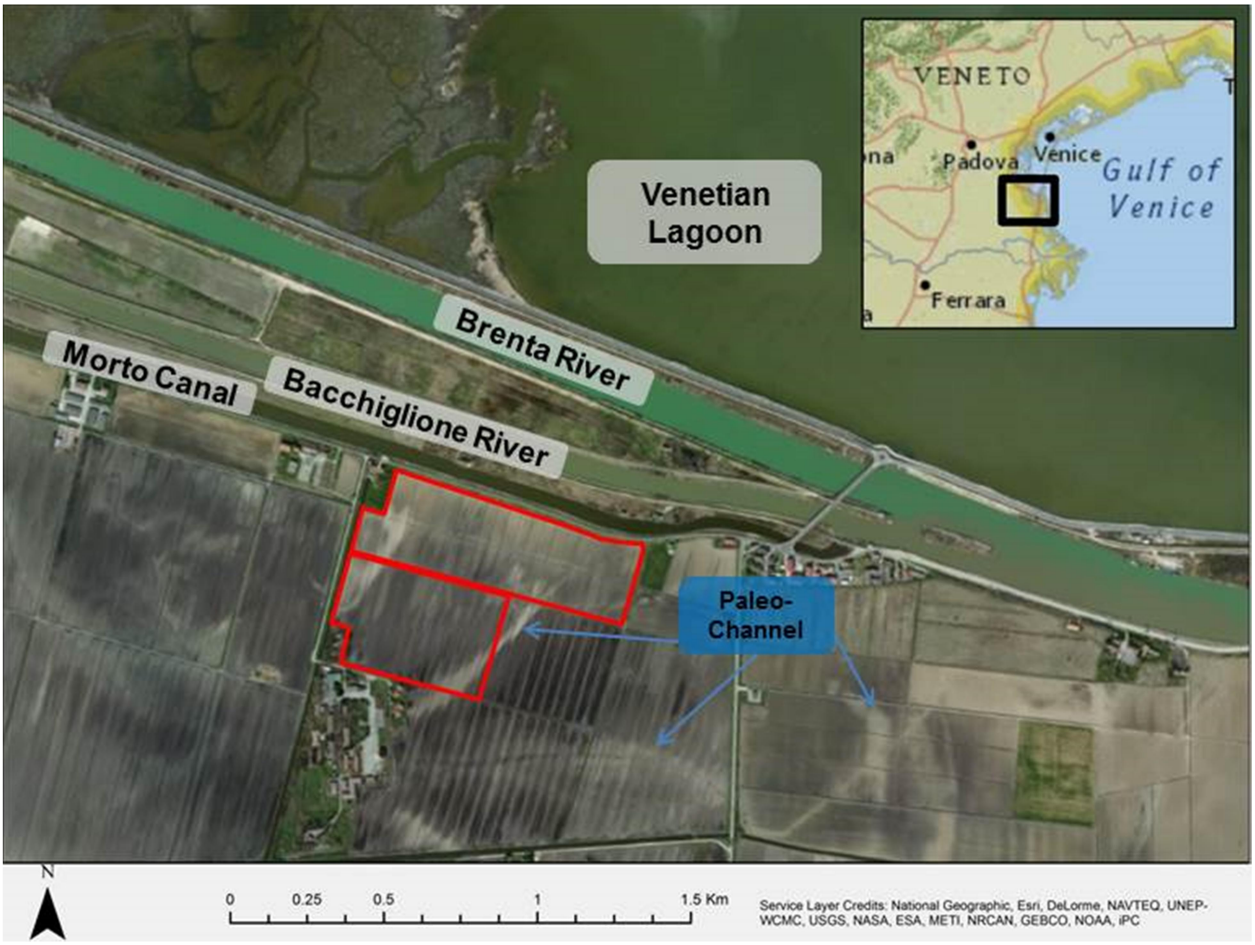


Figure 1.4. Map of the Study site which is located on the southern edge of the Venetian Lagoon. The red rectangles are the study site boundaries. The site is known to have paleo-channels (marked in blue) which can be verified in the aerial photography.

The current work will partially address points 2, 3, and 5 from the overall description. Specifically the aim of this research is to investigate the spatial distribution of soil hydraulic properties, understand the relationships between these properties and the soil geologic structure and to develop a prediction method which can be used as an aid in upscaling models to accomplish the tasks described in point 5 of the overall project description. These tasks will be accomplished through the use of multivariate geospatial statistics. Partial funding for the current work was provided by the USDA-NIFA International Science and Education Program, Award Number 2010-5160-2150 through the TransAtlantic Precision Agriculture Consortium (TAPAC). Contributions were also made by the research

program titled GEO-RISKS: Geological, Morphological and Hydrological Processes: Monitoring, Modeling, and Impact in the North-Eastern Italy, WP4, University of Padova, Italy.

Citations

Assouline, S. and D. Tartakovsky. 2001. Unsaturated hydraulic conductivity function based on a soil fragmentation process. *Water Resources Research*. 37(5): 1309-1312.

Brady, N. and R. Weil. 2008. *The Nature and Properties of Soils*. 14th ed. Upper Saddle River, NJ: Pearson.

Brooks, R. and A. Corey. 1964. Hydraulic Properties of Porous Media. *Colorado State University Hydrology Paper 3*. 27.

Burdine, N..1953. Relative permeability calculations from pore size distribution data. *Petroleum Transactions, American Institute of Mining Engineering*. 198: 71-77.

Carbognin, L., P. Teatini, L. Tosi. 2005. Land Subsidence in the Venetian area: known and recent aspects. *Giornale di Geologia Applicata*. 1: 5-11.

Chilés, J. P. and P. Delfiner. 2012. *Geostatistics: Modeling Spatial Uncertainty 2nd ed*. Hoboken, NJ: John Wiley & Sons, Inc.

Corwin, D. and S. Lesch. 2005. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*. 46: 11-43.

Custodio, E. and G. Bruggeman. 1987. Groundwater problems in coastal areas. *Publication of the IHP Working Group on Changes in the salt-fresh water balance in deltas*, *estuaries and coastal zones due to structural works and groundwater exploitation. 45-596.*

Durner, W.. 1994. Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*. 30(2): 211-223.

Gambolati, G., M. Putti, P. Teatini, and G. Gasparetto Stori. 2006. Subsidence due to peat oxidation and impact on drainage infrastructures in farmland catchment south of Venice Lagoon. *Environmental Geology*. 49: 814-820.

Goovaerts, P.. 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*. 43: 597-619.

Groovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physic-chemical soil properties. *Biol Fertil Soils*. 27: 315-334.

Groovaerts, P.. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*. 89: 1-45.

Manzini, G. and S. Ferraris. Mass-conservative finite volume methods on 2-D unstructured grids for the Richards' equation. *Advances in Water Resources*. 27: 1199-1215.

Mualem, Y.. 1976. A New Model for Predicting the Hydraulic Conductivity of Unsaturated Conductivity of Unsaturated Porous Media. *Water Resources Research*. 12(3): 513-522

Radcliffe, D. and J. Simunek. 2010. *Soil Physics with HYDRUS: Modeling and Applications*. Boca Raton, FL: Taylor and Francis Group.

Richards, L..1931/2004 Capillary conduction of liquids through porous mediums. Journal of Applied Physics. 1: 318-333. doi: 10.1063/1.1745010.

Rivoirard, J.. 2001. Which Models for Collocated Cokriging?. Mathematical Geology. 33(2): 117-131.

Ross, P., J. Williams, and K. Bristow. Equation for Extending Water-Retention Curves to Dryness. *Soil Science Society of America Journal*. 55: 923-927.

Simunek, J., O. Wendroth, and M. Th. van Genuchten. 1998. Parameter Estimation Analysis of the Evaporation Method for Determining Soil Hydraulic Properties. *Soil Science Society of America Journal*. 62: 894-905.

Simunek, J., M. Sejna, H. Saito, M. Sakai, and M. Th. van Genuchten, The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.0x, *Hydrus Series 3,* Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA, 2008.

Tobler W.. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(2): 234-240.

van Genuchten, M.. 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soil. *Soil Science Society of America Journal*. 44(5): 892-898.

van Genuchten, M. Th.. 1982. A comparison of numerical solutions of one dimensional unsaturated-saturated flow and mass transport. *Advances in Water Resources*. 5: 47-55.

van Genuchten, M. Th. and D. Nielsen. 1985. On describing and predicting the hydraulic properties of unsaturates soils. *Annales Geophysicae*. 3(5): 615-628.

Wackernagel, H.. 2003. *Multivariate Geostatistics 3rd ed*. Berlin, Heidelberg, Germany: Springer-Verlag.

Chapter 2: Explanation of the Spatial Variation of Hydraulic Properties of Soils in the Venice Lagoon Margin (Italy) with Factorial Kriging Analysis[1][2]

---

[1] To be submitted to *Geoderma*

[2] Authors: James Bevington, Francesco Morari and George Vellidis

Abstract

Understanding water movement in the vadose zone is critical for accurate climate and crop modeling, precision agriculture, soil-atmosphere gas exchanges, and contamination mitigation. The size and resolution of current studies is hindered by lab measurements of hydraulic properties because they are time consuming and costly. One reason that hydraulic measurements are difficult is because they are scale dependent due to the inherent heterogeneity of the soil. A better understanding of the spatial variability and underlying processes responsible for this variability is needed. Factorial kriging analysis (FKA) is a geospatial technique which identifies spatial relationships and common sources of variability. FKA was applied at 4 depths to a 20.8 ha field in the Po River delta which in known to have paleo-channel structures and highly heterogeneous soil. Texture, bulk density, and the van Genuchten-Mualem parameters K_s (saturated conductivity), α (inverse of air entry), n (shape parameter), θ_r (residual water content), and θ_S (saturated water content) were included in the analysis. Two nested spherical models with ranges around 105 m and 235 m plus nugget fit the experimental variograms and cross-variograms best. Regionalized correlation coefficients and regionalized PCA revealed many strong relationships which were not obvious from descriptive statistics. One example is the effect of interaction between texture and bulk density on n and K_s. Another is that K_s is more closely related to bulk density (BD) than texture. The first principal components (PCs) of the regionalized PCA explained the majority of the variability and the second PCs were rarely informative. The spatial distributions of the first PCs resembled bulk density at short scale and the paleo-channels and texture at long scale. The influence of the short scale PCs are greater than the long scale PCs near the surface but becomes less as depth increases. This suggests that depth plays an important role in the causal processes and should be considered more often. Future research should investigate spatial relationships between hydraulic parameters and other variables and attempt to use geospatial statistics as a tool to predict hydraulic properties.

List of Abbreviations: BD, bulk density; ECa, apparent electrical conductivity; FKA, factorial kriging analysis; K_s, saturated conductivity; *l*, connectivity and tortuosity parameter; *n*, shape parameter; NDVI, Normalized Difference Vegetation Index; PCA, principal component analysis; PTF, pedotransfer function; SOC, soil organic carbon; UHC, unsaturated hydraulic conductivity curve; WRC, water retention curve; α, inverse of air entry; θ_r, residual water content; θ_s, saturated water content;

1. Introduction

The vadose zone is of utmost importance in solving some of today's most challenging issues. In particular, this zone is an integral part of the hydrological cycle and the interface between groundwater, vegetation, and atmospheric interactions. These interactions are critical in understanding climate and increasing agricultural efficiency, two highly active areas of research. Water movement in the vadose zone differs greatly from water movement in other systems. Though much research on the complex interactions of soil and water movement has been completed, there is much progress to be made (Vereecken et al., 2010).

Unsaturated water retention curves (WRC) and hydraulic conductivity curves (UHC) are most often described by the van Genuchten model (1980) and Mualem model (1976). Several studies have concluded that lab measurements differ from field measurements of the hydraulic parameters of the WRC and UHC (Wierenga et al., 1991; Mallants et al., 1997b). Spatial heterogeneity is suggested as the primary source of error. In particular, Mallants et al., 1997b showed this to be true for measurements of saturated hydraulic conductivity (K_s).The variance of the K_s measurements decreased as sample volume increased, likely due to the existence of macropores and preferential flow paths. Several studies have evaluated short scale variability (<25 m) of hydraulic parameters (Ciollaro and Romano 1995; Mallants et al., 1997a; Hammel et al., 1999; Russo and Jury, 1987; Deurer et al., 2000;), while most larger scale studies have focused on water content (θ) instead of hydraulic parameters (Botros et al. 2009; Bocchi et al., 2000; Goulard and Voltz, 1992; Voltz and Goulard, 1994). However, few have evaluated large scale spatial variability of hydraulic parameters. Spatial cross-correlation of hydraulic parameters has also

been investigated mostly at short scale but never in a comprehensive multivariate geospatial framework (Ünlü et al., 1990; Vauclin et al., 1994; Mallants et al., 1996; Mertens et al., 2002; Greminger et al., 1985)

Development of pedotransfer functions (PTFs) to predict hydraulic parameters from soil properties is an active area of research. These studies often incorporate advanced multivariate statistical techniques to build their prediction models. Inclusion of new and different explanatory variables is also an active are of research (Vereecken et al., 2010). Doussan and Ruy (2009) used apparent soil electrical conductivity (ECa) to predict UHC. This is certainly a novel approach; moreover, electrical surveys can also hold tremendous information about the spatial variability of soil (Corwin and Lesch, 2005). The major drawback to measuring unsaturated hydraulic parameters is the time and cost involved and ideally would be reserved for cases when accuracy takes priority, e.g. shrink-swell studies. PTFs are best applied when accuracy can be sacrificed in exchange for sample quantity e.g. when studying water movement of a field or catchment. Also, PTFs are most applicable in the region where the data were collected but also depend on the scale of the collected data. Therefore, a PTF that is valid for a large area is unlikely to model small scale phenomena well because of the difficulty in measuring many samples (Vereecken et al., 2010). Applications of downscaling to these widely applicable PTFs could benefit from knowledge of the spatial structure of the variables involved. Regardless of the PTF's use, development of PTFs could benefit from a deeper understanding of causal processes and new insight into relationships between soil variables and hydraulic parameters. Factorial kriging analysis (FKA) offers this opportunity (Goovaerts, 1998).

FKA is a geospatial technique which emphasizes scale as part of relationship exploration. One of the primary tasks is to separate variance according to spatial scales. At each scale, relationships are examined and common sources of variance in the spatial distributions identified. Sollitto et al. (2010)

and Nanos and Martin (2012) were able to distinguish anthropogenic from natural sources of variation in trace elements typically associated with pollution. Dobermann et al. (1995) used FKA to investigate the relationships between soil chemistry, leaching, and hillslope position. Bocchi et al. (2000) applied FKA to soil properties and water content at several potentials. They explored the relationships between soil properties and water retention at different potentials and identified the effects of manure spreading on water retention. The only published use of FKA in conjunction with hydraulic parameters and soil properties is Biswas and Si (2009). They studied a 384 m 1D transect in an area characterized by glacio-fluvial and fluvial lacustrine derived sandy loams. While they did explore novel relationships between hydraulic parameters, texture, bulk density and SOC, they did not include depth. No underlying process was identified for the spatial distributions either.

Little scientific literature is available describing efforts to understand the underlying processes responsible for the patterns of variability seen in soil and hydraulic data. The goal of this study was to develop such techniques and apply them to a highly heterogeneous field in the Venice Lagoon Margin. The specific objectives of this study were to:

1. Explore field scale spatial relationships between hydraulic parameters;
2. Explore common sources of variability between hydraulic parameters and soil variables; and
3. Identify underlying processes responsible for spatial distributions found in the study

2. Methods

2.1 Study Site Description

The study site is a 26.5 ha field located on the southern shores of the Venetian Lagoon (figure 2.1). Due to wet conditions, a portion of the study site was abandoned leaving only 20.8 ha. This area was reclaimed from the lagoon by the construction of levees between 1892 and 1967 (Gambolati et al., 2006). A pumping station and a dense network of ditches control the depth to the water table, which is

generally maintained at ca. 0.6 m during the summer season in order to promote sub-irrigation. On the northern edge are the Morto Canal, Bacchiglione River, Brenta River and Venetian Lagoon. All of the canals and rivers are not natural as the rivers were diverted here at various times in the past. During construction of the Morto Canal, soil was placed on the northern edge of the field burying the original surface. Because of its reclaimed origin, soil shows an high variability in terms of texture, pH, salinity and presence of organic horizons, which can be superficial as well as buried a deeper depth. Two main soil typologies can be identified according to USDA (1998). In the northern part, along the Morto Canal, soils are Fluvaquentic Endoaquepts fine-silty, mixed, calcareous, mesic, and in the southern part Typic Sulfisaprists euic, mesic. Moreover, there are well preserved sandy paleo-river channels which cut through the field. These are artifacts of its geological history as a river delta. In recent years, the field has been continuously used to grow corn (*Zea mais* L.) and is tilled to a depth of approximately 30 cm each year.

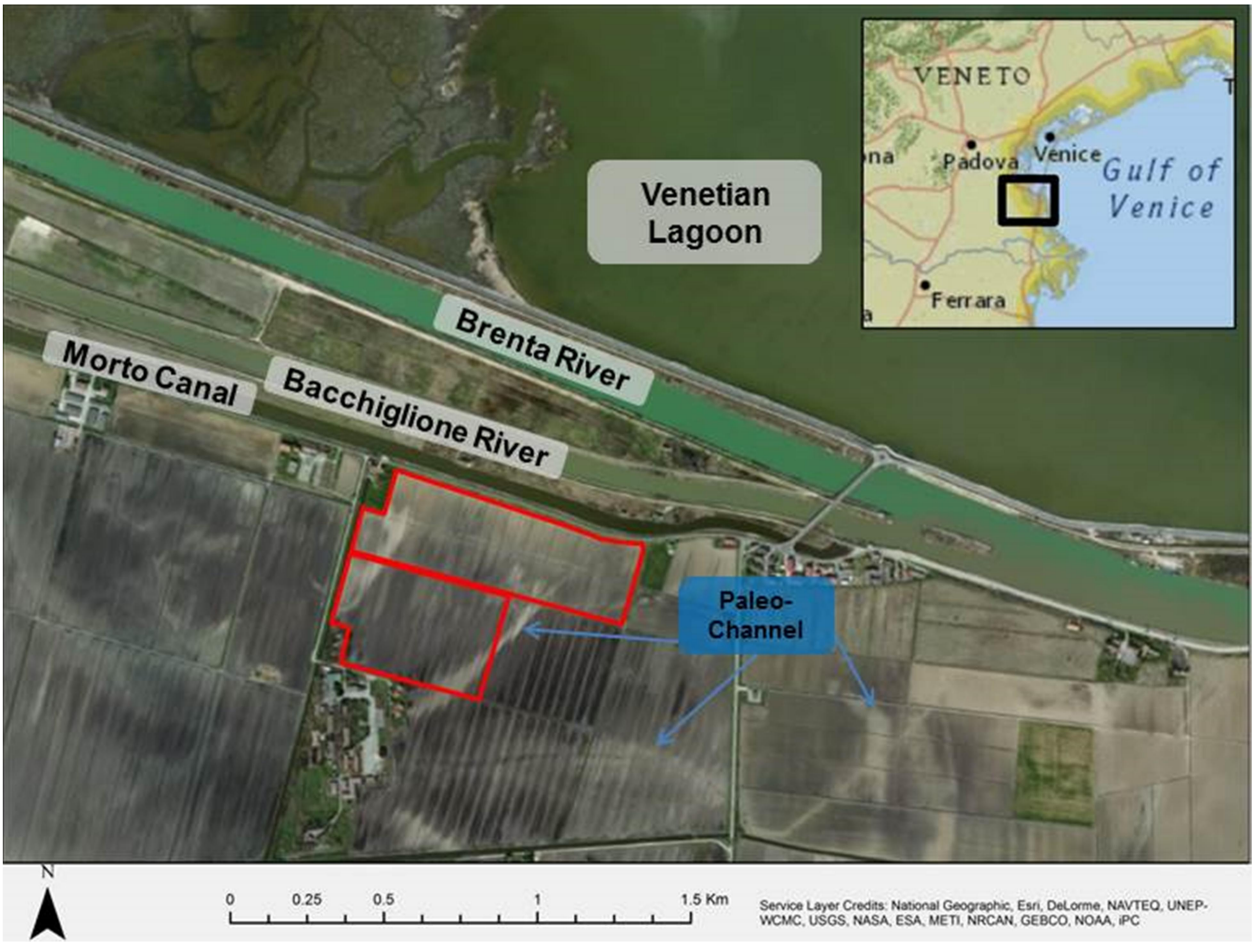


Figure 2.1. Map of the Study site which is located on the southern edge of the Venetian Lagoon. The red rectangles are the study site boundaries. The site is known to have paleo-channels (marked in blue) which can be verified in the aerial photography.

2.2 Secondary Variables

Exhaustive data sets are useful to geospatial studies when used as secondary variables and can greatly improve the prediction of primary variables (Wackernagel, 2003). For this study, 4 exhaustive data sets were collected, ECa at 0-0.75 m (ECa075) and 0-1.5 m (ECa150) depths, elevation, and bare soil reflectance as measured with the Normalized Difference Vegetation Index (NDVI). ECa datasets were collected in April 2011 using a CMD-1 frequency-domain electric induction sensor (GF Instruments, Brno, Czech Republic). The CMD-1 only collects one depth a time because it must be reconfigured for different depths. The data logger collects two measurements at each point; all values at the same location were

averaged together to have one representative value for the point. For ECa075, 18,053 measurements were collected at 9,017 locations. 20471 ECa150 measurements were collected at 10,221 locations. Elevation was measured in February 2012 using a Trimble FM 1000 CNH GPS receiver with real time kinematic (RTK) differential correction (Trimble Navigation Ltd., Sunnyvale, CA, USA) at 1,564 locations. Bare soil NDVI was measured in March of 2012 using an APS1-CropCircle (Holland Scientific, Lincoln, NE, USA) at 10,214 locations. This sensor uses light reflectance at 590 nm (VIS) and 880 nm (NIR) to calculate NDVI using $NDVI = (NIR - VIS)/(NIR + VIS)$ (Rouse et al., 1973). Each dataset was transformed, de-trended and interpolated using ordinary kriging to a 1 m resolution grid of 550 by 750 nodes using ISATIS v13.01 (Geovariances and Ecole des Mines de Paris, Avon Cedex, France) and the techniques described below. Interpolated values at the grid nodes were exported to Matlab (MathWorks, Natick, MA) for a principal component analysis (PCA). The loading of each node on principal component 1 (F1) was then used as a secondary variable in the geospatial study.

2.3 Soil Sampling

There are 123 sample locations which were selected under direction from an ECa survey as described in Scudiero et al. (2011). Disturbed samples were collected at all 123 locations in May 2010 by means of a motorized auger and undisturbed samples were taken at 50 of the 123 locations in October 2010 using a hydraulic core sampler. At each disturbed location, samples were collected at 3 depths: 0-0.15 m, 0.15-0.45 m, and 0.45-0.80 m. At each undisturbed location, samples were taken at 4 depths: 0.05-0.11 m, 0.25-0.31 m, 0.45-0.51 m, and 0.65-0.71 m. Texture was measured on all disturbed samples using a Mastersizer 2000 (Malvern Instruments Ltd, Great Malvern, UK) while soil organic carbon (SOC) was measured at 41 locations using a vario MACRO elemental analyzer device in CNS mode (Elementar, Hanau, Germany). Hydraulic measurements were made on undisturbed samples followed by bulk density measurements. The Core method (Grossman and Reinsch, 2002) was used to make bulk density measurements.

Hydraulic measurements were made on a Ku-pF Apparatus DT 04-01 (Umwelt-Geräte-Technik GmbH, Müncheberg, Germany) using the Wind method (Dane and Topp, 1994). The Wind method is an evaporation technique that starts with a saturated soil sample and allows water to evaporate from only the top surface. At predefined time intervals, tension is measured at two depths as well as change in mass. A fan aids in establishing a mostly stable gradient in the sample at the beginning of the experiment. An undisturbed soil sample with a length of 6.1 cm and circular cross-section of 41 cm^2 was used. Tensiometers with a range of 0-90 kPa and resolution of 0.01 kPa were placed at 1.5 cm and 4.5 cm from the top surface. The scale used had a resolution of 0.01 g. All instruments were read and recorded on a 10 min interval. Experiments were stopped when the top tensiometer reached a value of about 80-90 kPa or when tension data were not reliable (e.g. bottom value higher than the top one). At the end of the analyses, the gravimetric water content was determined after 12 h in an oven at 105° C.

From these measurements, water content and a flux could be calculated. However, calculating water content involves some assumptions such as a linear tension gradient (Operating Instructions for Ku-pF Apparatus DT 04-01). To avoid this, an inversion method was desirable. To improve inversion results and reduce the number of data points to be inverted, an averaging scheme was employed. Points were binned on a $\log_2$ scale of tension. Mass, upper tension, lower tension and time were each averaged to make a representative point for each bin. A $\log_2$ scale was used to preserve the rapid changes that happen early in the experiment and reserve the bulk of the data volume reduction for the end of the experiment where changes happen more slowly. Inversion of these representative points was done using HYDRUS 1D v4.12 (Simunek et al., 2008) and yielded the van Genuchten and Mualem parameters used for the study. Hydrus-1D is typically used for 1-dimensional analysis of water flow and solute transport in variably saturated porous media. Hydrus-1D is a mass-lumped linear finite elements code that uses evaporation flux and initial guess hydraulic parameters to predict tension values. The predicted values are compared to the measured values by evaluating an objective function. The

objective function is minimized using the minimized using the Marquardt-Levenberg scheme which provides an improved guess for the next iteration. Iterations continue until a stopping criterion is met. For this study, water content tolerance of 0.001% and pressure head tolerance of 1 cm of head was used.

2.4 Spatial Statistics

It is well documented in the literature that hydraulic parameters tend to not have normal distributions (Mallants et al., 1997b; Mallants et al. 1996; Bostros et al. 2009). Though it is not an explicit requirement of kriging, extreme distributions may yield unreliable results (Alary and Demougeot-Renard, 2010). For this reason, all variables were transformed using a Hermitian transformation algorithm (Wackernagel, 2003). After transformation, all variables were de-trended by fitting a constant, first or second order polynomial model to the data to satisfy the requirement of stationarity. Quality of fit was evaluated by calculating the standard deviation of the residuals. The model that produced the least variance in the residuals was selected. The four depths were treated as independent data sets. Most of the geostatistical analysis was performed using ISATIS. However, some tools from ArcMap 10.1 (ESRI, Redlands, CA, USA) were also used.

2.5 Variography

Experimental variograms and cross-covariograms were calculated. Because of the semi-discontinuous interface between paleo-channels and surrounding material and the arrangement of data locations, 1 to 2 points may be masked during this calculation. Bin sizes were selected by visual inspection of a bar graph depicting the number pairs in each bin. The bin size that resulted in the most even distribution and remained within a range of reasonable parameters was selected. Bins with relatively few points were removed from further analysis as to not give these few points too much

influence over the rest of the data. Bins with less than 5 pairs were removed which typically happened at the shortest h lag.

Many models were fit to the experimental variograms and cross-covariograms. Quality of fit was evaluated by calculating the sum of squared residuals, Akaike criterion and Bayesian Information criterion. The linear model of coregionalization LMC was used as a constraint in all multivariate models to ensure that the matrix of variogram functions was negative semi-definite (Goovaerts, 1992). This model assumes that all variables are influenced by the same processes and therefore uses the same set of ranges for all variograms and cross-covariograms (Wackernagel, 2003). Nested models with up to three structures were tested (including nugget effect). Models were fit using ISATIS's gradient method optimization algorithm (ISTATIS Help File, Geovariances and Ecole des Mines de Paris, 2013). On occasion, manual modifications to the model were made if the algorithm resulted in unreasonable results such as ranges well beyond the experimental variogram or at extremely short distances where there were no data points.

2.6 Factorial Kriging Analysis

Because FKA has been thoroughly described in the literature (Goovaerts, 1992; Chilés and Delfiner, 2012; and Wackernagel, 2003), only a brief overview is given here. FKA consists of three steps: fitting multivariate variogram and cross-covariogram models, calculating a regionalized PCA, then mapping the regionalized PCs. The mapped PCs visually show the sources of variance and can be interpreted to infer the underlying processes. After a model was fitted, regionalized correlation coefficients were calculated from the respective sill values. Using the regionalized correlation matrices as input, a regionalized PCA was computed for each range as a normal PCA would be. Maps of each PC were made by interpolating the projection of the measured points on the PC of interest. Secondary data

were incorporated in a multi-collocated cokriging scheme to improve accuracy of interpolation and therefore also the maps.

3. Results and Discussion

3.1 Descriptive Statistics

Table 2.1 gives a summary of the primary variables used in this study. One interesting observation is the mean of *l*, the connectivity and tortuosity parameter, is very near -1. Recently, Schaap and Leij (2000) found that -1 was a better value for *l* than 0.5 which is a commonly used value suggested by Mualem (1976). For the most part, the distributions are not normal. *l*, K_s, *n*, and α have strongly skewed distributions. Sand, silt, clay, and θ_r have slightly skewed distributions. Bulk density and θ_s have a bimodal distribution which is most obvious on the lower range of bulk density as seen in figure 2.2C. One hypothesis for this is the peat soils known to be in the area (Gambolati et al., 2006). If this is the reason, a similar bimodal distribution of SOC would be expected as seen with θ_s in figure 2.2A. A visual inspection of the SOC histogram (figure 2.2B) is indeterminate. A scatter plot of SOC vs BD (figure 2.2D) reveals a strong trend at the 8 cm and 28 cm depths and no trend or an unclear trend at 48 cm and 68 cm depths. This shallow versus deeper grouping was common amongst many variables. A likely explanation is the annual tillage to a depth of approximately 30 cm (Scudiero et al., 2013). In general, the layers above 30 cm were very different from those below 30 cm.

Table 2.1. Summary statistics of primary variables

	Obs	Mean	Median	Min	Max	Std. Dev	CV	Skew	Kurt
BD (g/cm^3)	166	0.900	0.978	0.139	1.521	0.356	39.5	-0.60	-0.54
SOC (%)	128	10.039	7.809	0.097	26.986	6.858	68.3	0.58	-0.75
Clay (%)	163	15.73	15.06	0.85	34.26	8.25	52.4	0.38	-0.78
Sand (%)	163	45.39	43.78	12.95	92.09	18.70	41.2	0.37	-0.49
Silt (%)	163	38.87	40.46	7.06	62.51	11.45	29.4	-0.74	0.07
K_s (cm/d)	164	243.16	101.04	0.18	2630.10	405.24	166.7	3.89	18.43
I	164	-1.007	0.001	-19.496	6.201	2.774	-275.3	-2.90	13.70
α (1/cm)	164	0.046	0.032	0.003	0.184	0.036	77.9	1.36	1.47
n	164	1.45	1.22	1.07	4.14	0.53	36.9	2.68	8.27
θ_r (cm^3/ cm^3)	164	0.199	0.195	0.000	0.599	0.121	60.8	0.73	0.91
θ_s (cm^3/ cm^3)	164	0.603	0.580	0.298	0.950	0.129	21.4	0.63	-0.13

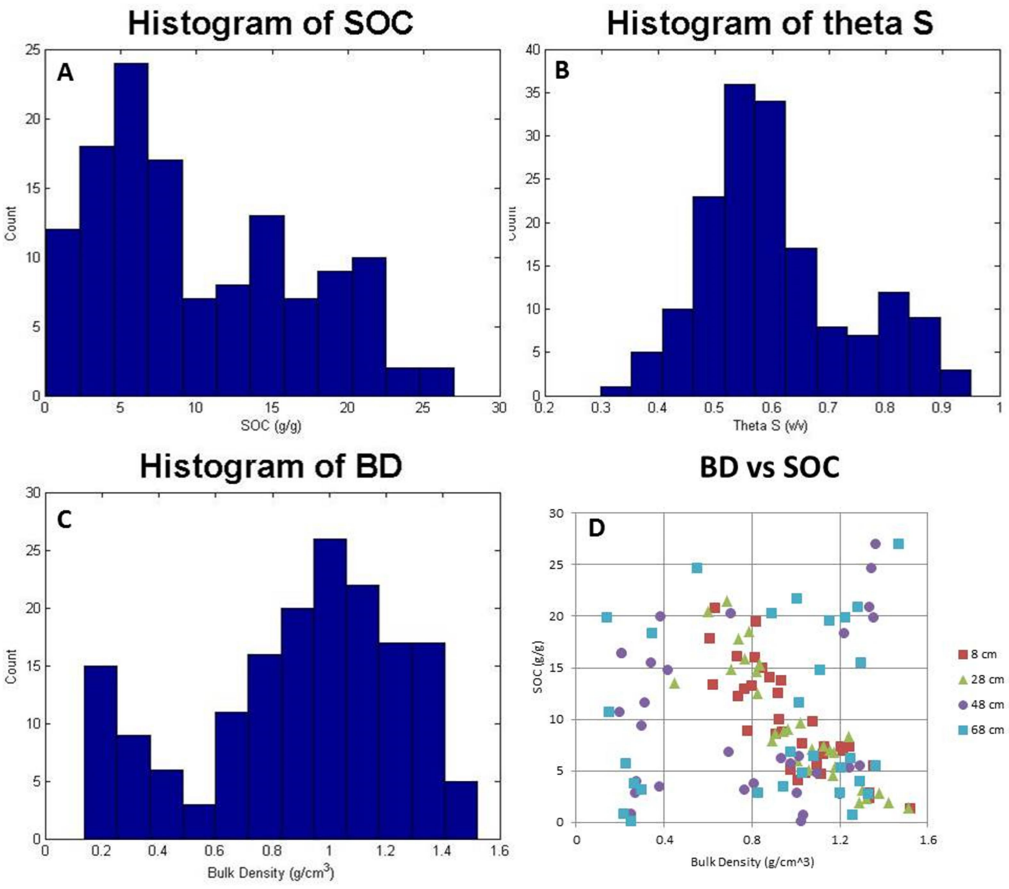


Figure 2.2A-D. A.) B.) C.) Histogram of SOC, θ_s, and bulk density respectivly. BD clearly has a bimodal distribution. This is also reflected in θ_s and to a lesser extent SOC. D.) Scatter plot of SOC vs BD grouped according to depth. 8 cm and 28 cm have a linear relationship while 48 cm and 68 cm are clusterd.

Another possible explanation for the bimodal BD distribution is soil structure. Structural information was not recorded on the soil samples. However, the hydraulic curves could hold some key information in the form of pore size distribution. A pore size distribution is approximated by the capacitance function which is the derivative of the WRC curve (Radcliffe and Simunek, 2010). Figure 2.3A shows a plot for all of the curves sorted by bulk density and figure 2.3B shows an averaged curve for each group. It appears that the high BD group may have two subgroups; one with typical mineral soil bulk density values and one with peat bulk density values. Regardless, it is clear that there is a trend between bulk density and the shape of the capacitance cure.

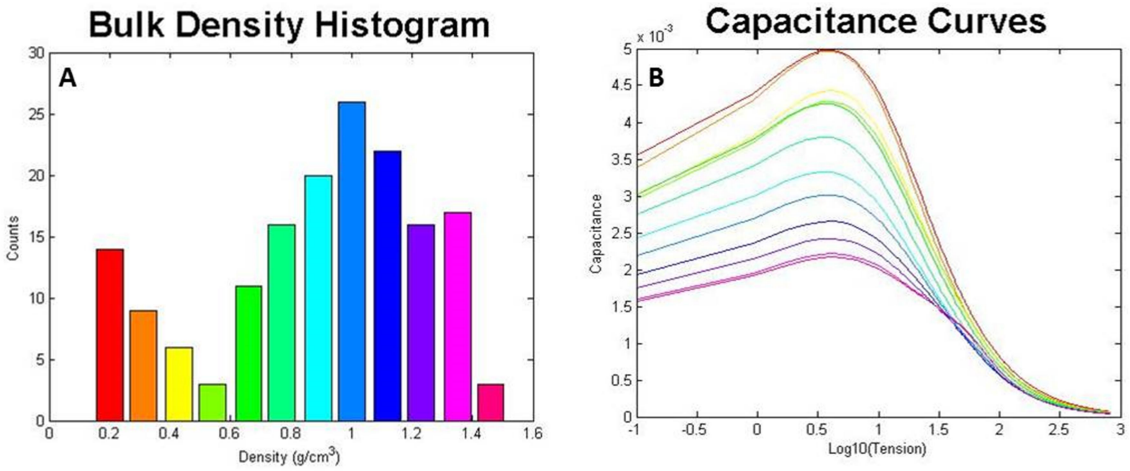


Figure 2.3. A.) Histogram of bulk density. The colors correspond to the curves in B. B.) Capacitance curves sorted by bulk density. Capacitance curves approximate the pore distribution. There is a trend of decreasing peak values with increasing bulk density.

All but one of the low BD samples (BD<0.5 g/cm^3) are located at 48cm and 68cm. Figure 2.4 reveals that these points are all near the northern edge and away from the paleo-channels. Because this front edge was filled in, deeper samples in this portion of the field will represent the original surface and will reflect properties of the original soils. The elevation change is reflected in the pattern.

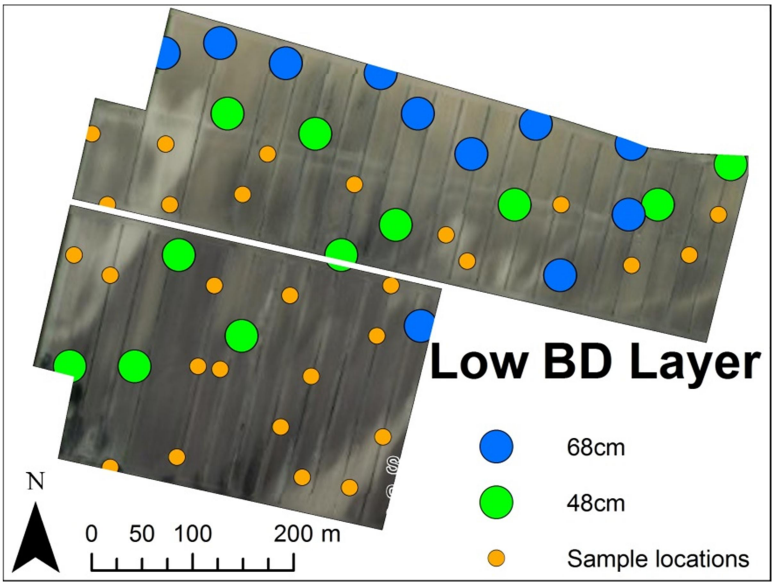


Figure 2.4. Map of the locations of low BD samples. The pattern reflects the field elevation pattern.

In order to add clarity to the analysis of relationships, all variables were transformed to a normal distribution with a mean of 0 and standard deviation near 1. Table 2.2 is a correlation table of the transformed variables calculated with casewise deletion. Transformed variables are denoted with an apostrophe. As expected, the texture variables, sand, silt, and clay are strongly correlated. Bulk density is not correlated with texture. This finding emphasizes that BD is a complex variable. Many of the well documented correlations among hydraulic variables are also observed, for instance α and K_s, BD and θ_S (Walczak et al., 2002; Botros et al., 2009; Mallants et al., 1996). One expected relationship that is not found is with texture and K_s.

Table 2.2. Correlation table of transformed variables calculated with casewise deletion

	BD'	SOC'	Clay'	Sand'	Silt'	K_s'	l'	α'	n'	θr'	θs'
BD'	1.000	-0.144	-0.075	0.129	-0.187	**-0.359**	-0.206	**-0.508**	0.182	**-0.334**	**-0.867**
SOC'	-0.144	1.000	-0.184	0.232	-0.212	0.027	-0.028	-0.020	0.130	0.057	0.121
Clay'	-0.075	-0.184	1.000	**-0.941**	**0.817**	0.005	-0.238	0.251	**-0.724**	-0.022	0.250
Sand'	0.129	0.232	**-0.941**	1.000	**-0.952**	0.033	0.246	-0.209	**0.702**	-0.034	**-0.318**
Silt'	-0.187	-0.212	**0.817**	**-0.952**	1.000	-0.057	-0.238	0.161	**-0.633**	0.109	**0.371**
K_s'	**-0.359**	0.027	0.005	0.033	-0.057	1.000	**0.410**	**0.587**	-0.144	-0.063	0.228
l'	-0.206	-0.028	-0.238	0.246	-0.238	**0.410**	1.000	-0.087	0.194	-0.043	0.107
α'	**-0.508**	-0.020	0.251	-0.209	0.161	**0.587**	-0.087	1.000	-0.221	0.279	**0.366**
n'	0.182	0.130	**-0.724**	**0.702**	**-0.633**	-0.144	0.194	-0.221	1.000	0.250	**-0.365**
θr'	**-0.334**	0.057	-0.022	-0.034	0.109	-0.063	-0.043	0.279	0.250	1.000	**0.352**
θs'	**-0.867**	0.121	0.250	**-0.318**	**0.371**	0.228	0.107	**0.366**	**-0.365**	**0.352**	1.000

(N=126) Bold font denotes significance of P<0.001.

3.2 Secondary Variables

A PCA of the residuals of the 4 transformed, exhaustive data sets, ECa075, ECa150, Elevation, and NDVI results are presented in table 2.3. Principal component 1 (F1) described 80.1% of the variance in the data. Both ECa data sets loaded strongly on this component and Elevation to a lesser extent. Elevation loaded in an opposite direction to ECa on F1. NDVI did not load on F1 in a significant way but loads strongly on Principal component 2 (F2). Because only F1 is used in subsequent analysis, NDVI does not add much to this study. From the map in figure 2.5H, it is clear that F1 represents the paleo-channels but there are also some other phenomena occurring.

Table 2.3. Summary of PCA results on exhaustive data

Parameter	F1	F2	F3	F4
Explained %	80.1	12.3	4.6	3.0
ECa075	0.6855	-0.1098	-0.3276	0.6409
ECa150	0.6399	-0.3091	0.5237	-0.4697
Elevation	-0.3294	-0.5807	0.5294	0.5235
NDVI	0.1097	0.7451	0.5815	0.3075

Each column represents a PC ordered by descending variance accounted for.

3.3 Spatial Statistics

3.3.1 Variography

Fitting the linear model of coregionalization (LMC) becomes more difficult as variables are added to the analysis because $n(n+1)/2$ variograms are required. For this reason, it is important to only retain variables that add value. Including three texture variables does not add new information because one is a linear combination of the other two (they sum to one). Sand and clay were used to represent texture because they represent the boundaries of the system. If one of these variables failed to add to the analysis, only one texture variable was included. *I* had very poor spatial structure and detracted from the quality of fit of the LMC so it was dropped and assumed to be a constant -1 in space. SOC at depths of 8 cm and 28 cm followed the same patterns as BD and added little new information. This is to be expected because of the strong global correlation. At depths of 48 cm and 68 cm, SOC had poor spatial structure characterized by high nugget effect. SOC was removed from the analysis. In every case, two nested spherical models fit the experimental variograms best.

Table 2.4 is a summary of the models. All models had three structures, nugget and two spherical. Clay was not used for 8 cm or 68 cm depths. F1 was used as a secondary variable for all depths except 8 cm where elevation was more informative. Over all, short ranges (short) tended to be around 105 m and long ranges (long) about 235 m. The quality of fit was assessed by the sum of squared residuals (SSR). Fit quality decreased with depth. This finding was both surprising and unexpected. Although the number of samples (N=4) used for this assessment is small, a linear model fits very well implying that the trend is strong (R^2=0.92).

Table 2.4. Summary of spatial models.

Depth (Nom.)	# of Vars[1]	Short[2] (m)	Long[2] (m)	SSR[3]	L/S[4]
8	9	106.8	224.5	0.028	0.759
28	8	153.0	285.4	0.049	0.877
48	9	74.2	209.9	0.094	1.627
68	8	92.0	216.5	0.098	4.267

1) # of Vars is the number of variables used in the spatial model.
2) Short and Long represent the short and long scale ranges for the spatial models respectively.
3) SSR is the sum of squared residuals of the experimental data to the fitted model.
4) L/S is the ratio of long to short scale Eigenvalues of the first PCs and represents a ratio of the variance represented by long scale processes to that of short scale processes.

3.3.2 Regionalized relationships

The regionalized correlations show many strong relationships that were not obvious from the descriptive statistics as seen in tables 2.5A-D. Finding these trends reveals the benefit of including scale and autocorrelation in soil analysis. The variables in this section are the transformed, de-trended residuals and are indicated with a *. In most cases, primary variables had a strong correlation with the secondary variable for at least one range. This means that secondary variables hold relevant information to this study.

Sand* and BD* show a positive correlation but not at 8 cm and 28 cm long range and 48 cm short range. Sand* and *n** have a very strong, positive relationship at long range; at short range, trends change with depth from strong positive at 8 cm and 28 cm to no correlation to negative correlation at 68 cm. This is interesting because sand' and *n*' were globally correlated. Typically variables that are globally correlated show the same correlation at both scales. In this case, the influence of the short scale process at the depths where the correlation coefficients are opposite is less than the influence of the long scale processes. See below for further explanation. Sand* and α* are negatively correlated throughout all ranges and depths, a finding opposite to many others (Biswas and Si, 2009; Li et al., 2006). The magnitude of the Sand* and α* relationships is surprising because globally, they were not very correlated. This indicates that the nugget effect on these variables masks important information. Except

for 8 cm long range, α* and BD* have a strong, negative correlation. Biswas and Si (2009) also found a negative α* and BD* relationship. The relationship between θ_s* and α* is similar to that of BD* and α*which makes sence because of the strong relationship between BD' and θ_s'. The only deviations are at 48 cm long scale and 68 cm short scale. In general, 48 cm long scale and 68 cm short scale have very different hydraulics than the other layers. BD* and K_s* and Sand* and K_s* are mostly negatively correlated except for at 48 cm long scale and 68 cm short scale. The relationships between BD* and K_s* are stronger than the relationships between texture* and K_s*. Often, when discussing K_s, texture is usually the first variable that comes to mind and one would usually expect a positive relationship (Jaynes and Tyler, 1984; Puckett et al., 1985) which is not the case here. These findings suggest that perhaps bulk density plays a larger role. BD* and *n** at short scale have a positive relationship at 8 cm and 28 cm and a negative one at 48 cm and 68 cm. At long scale, BD* and *n** have a positive relationship at all depths. At short scale, the BD* and K_s*relationship acts in the same direction as the BD* and *n** relationship at some depths, but at others, they act in different directions. Whether or not they act in the same or a different direction, appears to be related to the strength of the relationship between BD* and Sand*. If the BD*, Sand* relationship is strong, they act in opposite directions, if it is weak, they act in the same direction. This suggests that an interaction between texture and bulk density may influence relationships with these hydraulic parameters. K_s* and *n** are strongly, negatively correlated at 28 cm and 68 cm short scale and long scale. At 8 cm *n** and K_s* are not correlated at short scale and negatively correlated at long scale. At 48 cm, *n** and K_s* are positively correlated at short scale and not correlated at long scale. The short scale *n** and K_s* relationships reflect their relationships with BD*. α* and K_s* show a strong positive relationship at short scale except at 68 cm; at long scale 28 cm and 68 cm also show this relationship. It is difficult to determine whether the short or long scale process is more dominant in this relationship. At long scale, α* and *n** have a negative relationship with similar magnitudes to the relationship between α* and K_s*; at short scale, α* and *n** are negatively correlated

at 8 cm and 28 cm but become positively correlated at 48 cm and 68 cm. θ_r*, of all of the parameters tends to have the least physical significance and is viewed by many as a fitting parameter (Radcliffe and Simunek, 2010).

θ_r* also dominates the retention curves at high tensions beyond where data was collected. For this reason, θ_r* may be less reliable. θ_r* and θ_s* have a positive relationship at short scale except at 8cm and a positive relationship at 8 cm and 28 cm long scale. At short scale, θ_r* and *n** have very strong trends, negative at 8 cm and 28 cm, positive at 48 cm and 68 cm. At 8 cm long scale, θ_r* and *n** have a negative relationship which becomes positive at 28 cm and 48 cm. Again, global trends are lacking because of inconsistency in relationship direction. θ_r* and BD* have very similar trends to θ_r* and θ_s* only opposite in direction.

Many of the relationships are strong but they are inconsistent between depths. This inconsistency masks global trends and makes deciphering the processes difficult. One source of inconsistency is associated with the bimodal distribution of bulk density. The different modes likely act in very different ways. Perhaps separating the modes into a separate data set and studying them independently would be revealing. However, that is beyond the scope of this study. While many relationships are explored here, one potentially revealing variable is not part of this study, soil structure. Texture and bulk density have very inconsistent relationships with hydraulic parameters (Weynants et al., 2009; Schaap and Bouten, 1996; Wösten et al., 2001) and are not well understood. Soil structure is likely to capture the interaction effects between texture and bulk density (David Radcliffe, personal communication) which may aid in deciphering these complex relationships. Soil structure has also been suggested as a potential predictor variable for PTF (Pachepsky et al., 2006; Li et al., 2006).

Table 2.5A. 8 cm Regionalized correlation coefficients

	Elev*	**BD***	**Clay***	**Sand***	**K_s***	**α***	***n****	**θr***	**θs***
Elev*	1	0.892	-0.182	0.259	-0.954	-0.067	0.514	-0.785	-0.740
BD*	0.770	1	-0.011	0.023	-0.875	0.360	0.563	-0.803	-0.646
Clay*	-0.177	-0.549	1	-0.986	0.430	0.657	-0.779	0.600	0.767
Sand*	0.075	0.532	-0.988	1	-0.474	-0.751	0.727	-0.588	-0.777
K_s*	-0.722	-0.862	0.067	-0.077	1	0.122	-0.747	0.931	0.903
α*	-0.834	-0.931	0.674	-0.608	0.661	1	-0.115	0.049	0.321
n*	0.264	0.574	-0.994	0.964	-0.084	-0.726	1	-0.932	-0.928
θr*	-0.001	-0.092	0.827	-0.766	-0.425	0.378	-0.846	1	0.961
θs*	-0.698	-0.994	0.572	-0.571	0.856	0.900	-0.585	0.084	1

Bottom triangle is short range coefficients and top triangle is long range coefficients.

Table 2.5B. 28 cm Regionalized correlation coefficients

	F1	**BD***	**Sand***	**K_s***	**α***	***n****	**θr***	**θs***
F1	1	-0.999	-0.129	0.342	0.596	-0.312	0.437	0.856
BD*	-1.000	1	0.174	-0.374	-0.617	0.346	-0.390	-0.845
Sand*	-0.997	0.994	1	-0.916	-0.751	0.936	0.778	-0.144
K_s*	0.792	-0.810	-0.738	1	0.943	-0.999	-0.484	0.494
α*	0.994	-0.997	-0.982	0.853	1	-0.926	-0.175	0.754
n*	-0.800	0.817	0.747	-1.000	-0.860	1	0.530	-0.450
θr*	0.826	-0.842	-0.777	0.998	0.882	-0.999	1	0.509
θs*	0.953	-0.962	-0.924	0.940	0.980	-0.944	0.958	1

Bottom triangle is short range coefficients and top triangle is long range coefficients.

Table 2.5C. 48 cm Regionalized correlation coefficients

	F1	**BD***	**Clay***	**Sand***	**K_s***	**α***	***n****	**θr***	**θs***
F1	1	-0.568	0.699	-0.655	-0.398	0.486	-0.647	-0.335	0.567
BD*	-0.972	1	-0.672	0.618	0.535	-0.327	0.777	0.052	-0.955
Clay*	-0.084	0.206	1	-0.984	-0.030	0.288	-0.974	-0.759	0.682
Sand*	-0.069	0.245	0.777	1	-0.100	-0.314	0.972	0.777	-0.666
K_s*	0.553	-0.609	0.168	-0.379	1	0.175	0.072	-0.480	-0.472
α*	0.704	-0.772	0.054	-0.400	0.972	1	-0.289	-0.103	0.228
n*	0.968	-0.960	0.063	-0.072	0.712	0.834	1	0.636	-0.811
θr*	0.643	-0.628	0.397	-0.077	0.944	0.917	0.789	1	-0.067
θs*	0.988	-0.927	-0.051	0.034	0.455	0.606	0.930	0.585	1

Bottom triangle is short range coefficients and top triangle is long range coefficients.

Table 2.5D. 68 cm Regionalized correlation coefficients

	F1	BD*	Sand*	K_s*	α*	*n**	θr*	θs*
F1	1	-0.297	-0.618	0.031	0.357	-0.520	0.213	0.356
BD*	-0.992	1	0.826	-0.962	-0.924	0.885	0.164	-0.993
Sand*	-0.823	0.888	1	-0.666	-0.949	0.991	-0.357	-0.886
K_s*	-0.715	0.797	0.986	1	0.843	-0.757	-0.289	0.932
α*	0.087	-0.211	-0.637	-0.759	1	-0.983	0.226	0.961
n*	0.933	-0.971	-0.972	-0.919	0.439	1	-0.288	-0.935
θr*	0.607	-0.702	-0.951	-0.990	0.844	0.852	1	-0.048
θs*	0.998	-0.998	-0.859	-0.759	0.152	0.955	0.658	1

Bottom triangle is short range coefficients and top triangle is long range coefficients.

The results of the regionalized PCA are summarized in Tables 2.6A-D. At all depths and scales, the first PC explained the majority of the variance ranging from 94.4% to 61.0%. The second PC explained between 34.5% and 5.6% of the variance. Of the second PCs, only 28 cm long range was retained. The Eigen values quantify the variance represented by each PC in absolute terms. Because they are absolute, we can use them to compare short scale versus long scale PCs. In table 2.4, the column labeled L/S is the ratio of Eigen values for the first PC of long scale to short scale. This ratio indicates which process is more dominant. There is a strong trend between the L/S ratio and depth. At shallow depths, short scale processes are more dominant. As depth increases, long scale processes become more dominant.

At short scale, PC 1 (PC1S) is characterized by loadings of BD* and θ_s*. α* and K_s* also load on this PC except at 68 cm. As above, 68 cm deviates from the other layers at this scale. α* loads slightly stronger than K_s* at 8 cm and 28 cm and equally at 48 cm. This reflects the relative strengths in their relationships with BD*. After viewing the pore distributions in figure 2.3, it is not surprising that α* loads on the PC with BD*. Since there is a global trend and strong trends at this scale between K_s* and α*, it is also not surprising that K_s* loads too. As noted above, BD* appears to play a large role with K_s*. In this case, texture does not really load and the loading strengths over all are weak. This makes inferences

about a possible BD*, Sand* interaction impossible. θ_r* loads at depths of 48 cm and 68 cm, but not at 28 cm as might be expected given the strong relationships above.

At long scale, PC 1 (PC1L) is related to texture at all depths except 28 cm. At depths of 48 cm and 68 cm, *n** loads strongly on PC1L. 48 cm and 68 cm have similar behavior whereas 8 cm and 28 cm are more similar. The only difference between 8 cm and 28 cm is with interactions between BD* and texture*. At 8 cm texture* is dominant. However, there are two subsets where BD* acts in opposite ways. At 28 cm, it appears that BD* is partially dominant and texture* is partially dominant. However, it is important to note that the strongest loading on PC1L at 28 cm is F1. F1 and texture* did not correlate but BD* and F1 did. This PC likely illustrates that F1 is not representative of the data at this depth. This would not be a surprise because F1 did not represent 8 cm well and 8 cm and 28 cm tend to be similar. However, F1 correlated with more primary variables than Elevation*.

Table 2.6A. 8cm Regionalized PCA Summary

	Elev*	BD*	Clay*	Sand*	K_s*	α*	*n**	θr*	θs*	Eigen	% Var
PC 1S	0.107	0.436	-0.175	0.212	-0.300	-0.673	0.226	-0.065	-0.353	1.700	76.8
PC 2S	-0.105	-0.185	-0.348	0.434	0.547	-0.046	0.421	-0.384	0.137	0.431	19.5
PC 1L	0.197	0.220	-0.413	0.484	-0.201	-0.009	0.116	-0.400	-0.542	1.290	72.5
PC 2L	-0.335	-0.584	-0.431	0.483	0.207	-0.023	-0.012	0.263	0.128	0.458	25.7

Table 2.6B. 28cm Regionalized PCA Summary

	F1	BD*	Sand*	K_s*	α*	*n**	θr*	θs*	Eigen	% Var
PC 1S	-0.466	0.423	0.112	-0.318	-0.438	0.248	-0.118	-0.472	1.945	94.4
PC 2S	0.410	-0.319	-0.139	-0.619	0.188	0.468	-0.196	-0.186	0.116	5.6
PC 1L	0.745	-0.428	-0.145	0.169	0.141	-0.172	0.140	0.378	1.705	61.0
PC 2L	-0.165	0.068	-0.650	0.353	0.145	-0.394	-0.493	-0.052	0.965	34.5

Table 2.6C. 48cm Regionalized PCA Summary

	F1	BD*	Clay*	Sand*	K_s*	α*	*n**	θr*	θs*	Eigen	% Var
PC 1S	0.312	-0.462	0.005	-0.023	0.475	0.430	0.077	0.362	0.375	2.390	78.1
PC 2S	-0.332	0.407	0.075	-0.046	0.544	0.269	-0.041	0.301	-0.507	0.560	18.3
PC 1L	0.461	-0.151	0.453	-0.484	-0.009	0.104	-0.519	-0.120	0.178	3.888	77.9
PC 2L	0.811	-0.018	-0.203	0.283	-0.080	0.301	0.299	0.184	-0.012	0.622	12.5

Table 2.6D. 68cm Regionalized PCA Summary

	F1	BD*	Sand*	K_s*	α*	*n**	θr*	θs*	Eigen	% Var
PC 1S	-0.483	0.498	0.252	0.152	-0.076	-0.120	-0.539	-0.349	0.972	83.4
PC 2S	-0.445	0.302	-0.123	-0.138	0.332	-0.006	0.706	-0.263	0.193	16.6
PC 1L	-0.318	0.355	0.440	-0.216	-0.427	0.448	-0.041	-0.384	4.145	73.8
PC 2L	-0.860	-0.246	0.092	0.303	0.182	-0.025	-0.154	0.205	1.119	19.9

3.3.3 Spatial Distribution of PCs

In figures 2.5A-R, each retained PC is compared with maps of variables which have similar spatial distributions. Loadings may be negative and therefore maps may be reciprical yet have the same distribution. At 8 cm, the distribution of PC1S is similar that of BD*. The distribution of PC1L is similar to Elevation*. Elevation* is related to the paleo-channel. At 28 cm, the distribution of PC1S resembles BD* but deviates in the central portion of the field. PC1L is almost identical to F1, the secondary variable and PC2L resembles the distribution of Sand*. This confirms the suspicion that F1 does not represent this layer well at long scale. At 48 cm, the short scale PC1 map is identical to the map of BD* for this depth. The long scale PC at 48 cm also has a near identical match but to Sand*. This distribution resembles the shapes of the paleo-channels. The relationships at 68 cm are similar to those at 48 cm except there is a strong deviation in the southwest corner of PC1S from the distribution of BD*. Overall, the short scale distributions tend to resemble BD while the long scale distributions tend to resemble texture* and the paleo-channels.

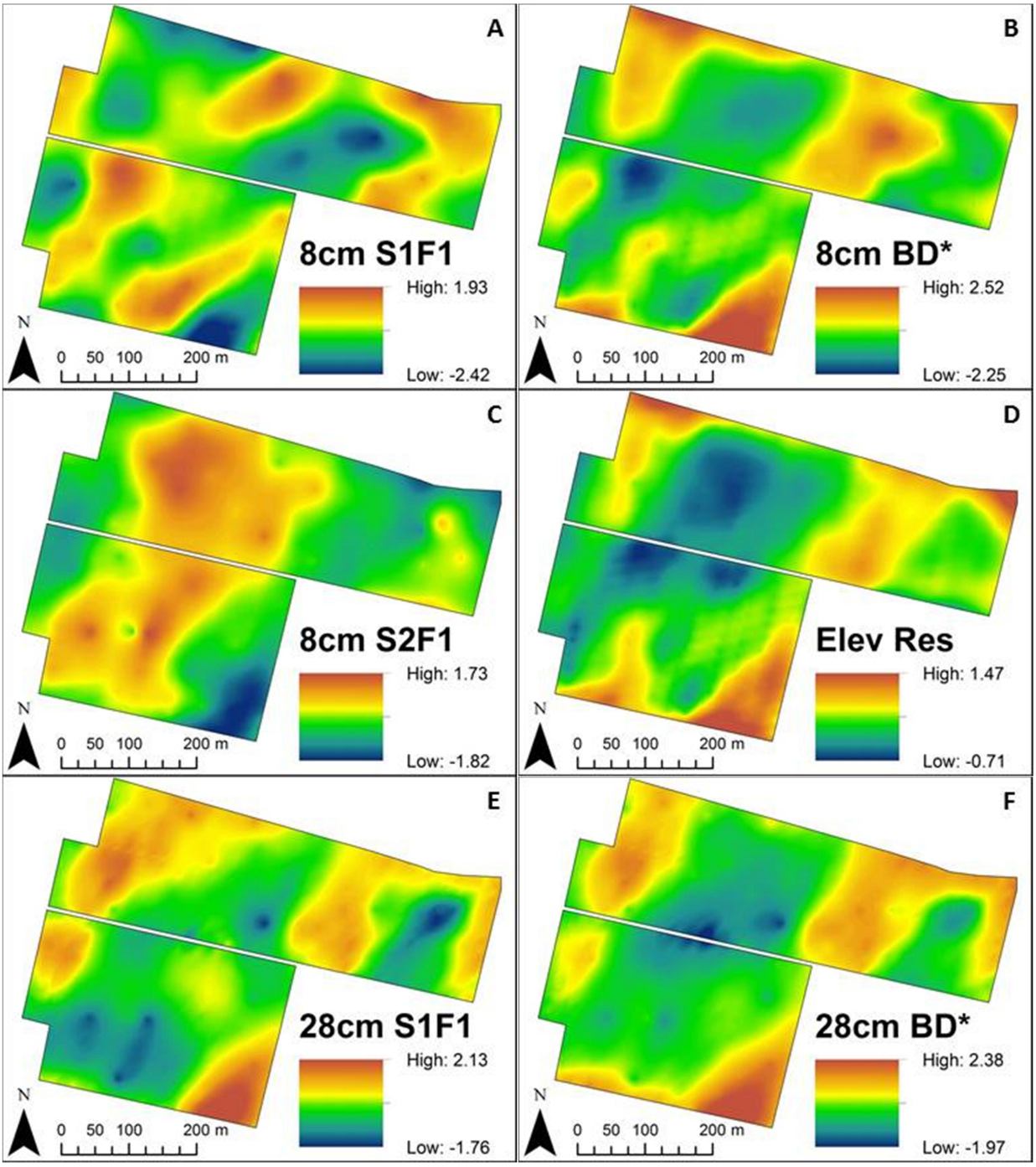


Figure 2.5A-F. Maps of retained PCs and their distribution counterparts. S1XX designates short range while S2XX designates long range. F1, F2, ect designates the principal component in ranked order from most variance explained to least variance explained. Maps in the left column are FKA results and maps in the right column are the best physical match to the FKA maps based on visual assessment of the spatial distribution.

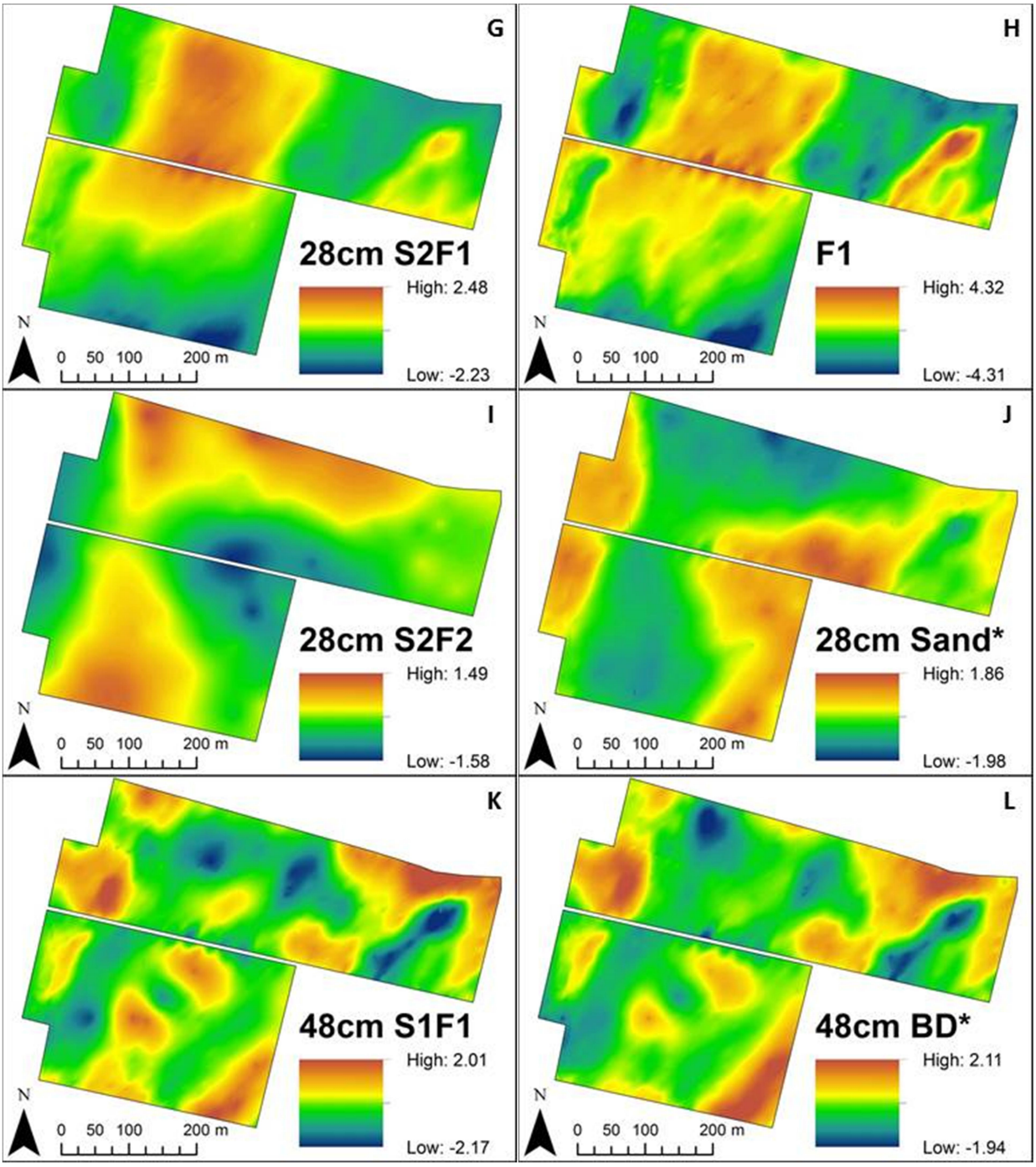


Figure 2.5G-L. Maps of retained PCs and their distribution counterparts. S1XX designates short range while S2XX designates long range. F1, F2, ect designates the principal component in ranked order from most variance explained to least variance explained. Maps in the left column are FKA results and maps in the right column are the best physical match to the FKA maps based on visual assessment of the spatial distribution.

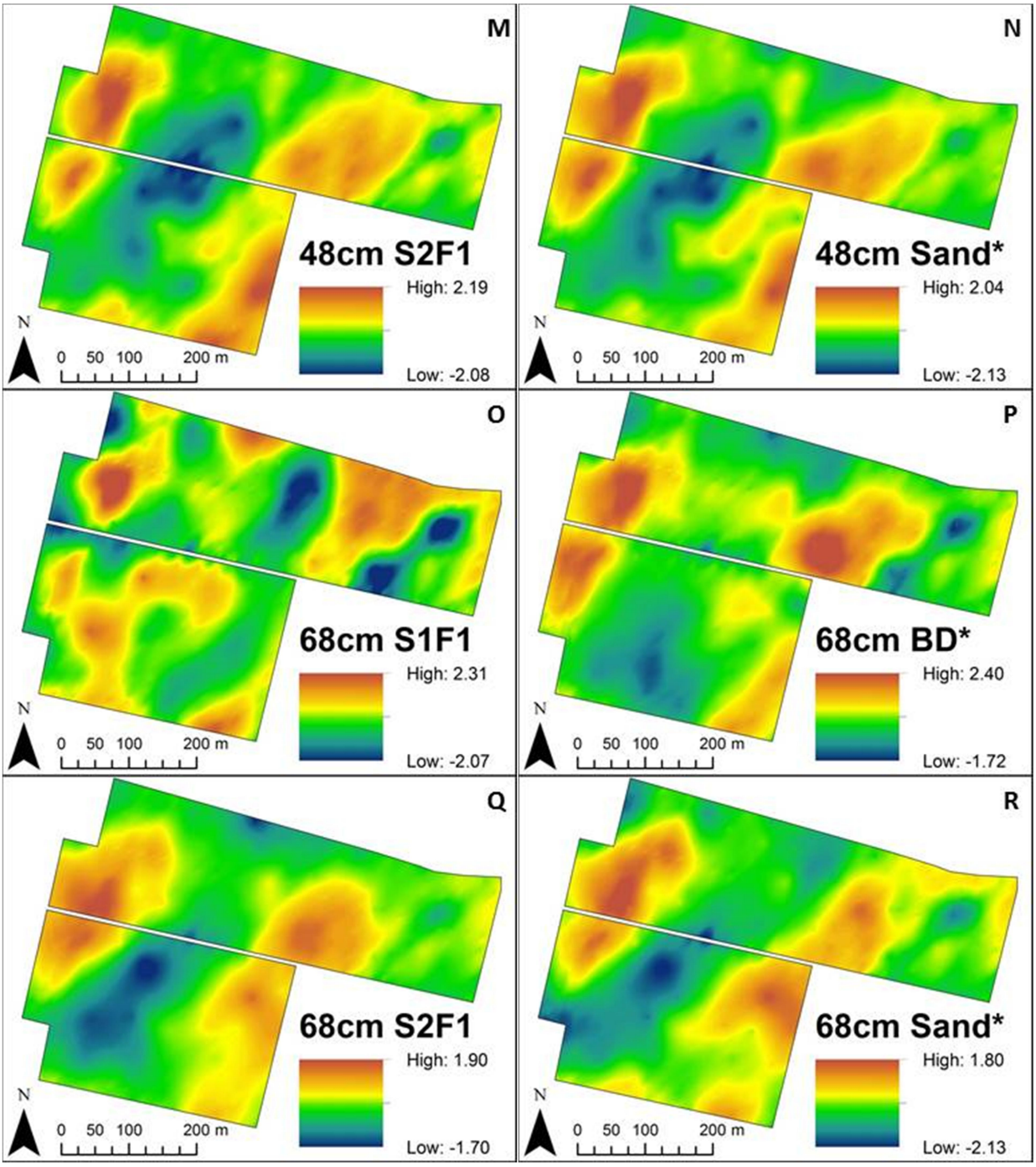


Figure 2.5M-R. Maps of retained PCs and their distribution counterparts. S1XX designates short range while S2XX designates long range. F1, F2, ect designates the principal component in ranked order from most variance explained to least variance explained. Maps in the left column are FKA results and maps in the right column are the best physical match to the FKA maps based on visual assessment of the spatial distribution.

4. Conclusions

The application of geospatial statistics to soil science has the potential to address some of the field's most pressing issues. In particular, FKA is useful for identifying the underlying processes that affect the soil variables that are regularly studied. This type of analysis is effective at identifying relationships by removing scale dependency from the variables. When this is done, relationships become strong and more obvious. A subsequent variable reduction reduces the variable space into common factors which represent the related processes that act in the soil. Maps of these common factors can be used to identify the real world process. An issue of concern in unsaturated hydraulic research is that laboratory and field measures often are not in agreement. This is likely caused by two sources of error, laboratory measurement errors and errors resulting from heterogeneity in the field. This study assesses and demonstrates the use of FKA and related techniques to describe the heterogeneity of the field and identify how this heterogeneity affects the variables of interest.

An analysis of van Genuchten and Mualem hydraulic parameters in conjunction with soil variables such as texture and bulk density was conducted. Several relationships between soil variables and hydraulic parameters as well as between hydraulic parameters were developed. Some of these relationships, like θ_s and BD are already well documented, but some of them, like the complex K_s, n, BD, and texture relationship are newer. Future work should investigate this interaction both from a soil physics standpoint and more in-depth geospatially. Mapping the spatial distribution of regionalized PCs revealed that their distributions are very similar to the distributions of non-hydraulic parameters. At short scale, distributions tended to resemble bulk density while long scale distributions resembled texture and the paleo-channel.

This study suggests that there are many variables to considerer when investigating field versus lab error. For instance, depth is an under rated variable. We see in this study that relationships between

variables change markedly with depth. At the surface there are many more processes occurring such as high levels of microbial activity, erosion, and temperature changes, many of which are too small to capture in a spatial sense. Even though many of these processes are not represented in this study, the results show that short scale processes are more dominant than long scale processes closer the surface. As depth increases, there are less short scale processes and long scale processes become more dominant. Land use is another important factor in the behavior of soil variables; in this study, there are distinct differences between tilled and untilled soil.

Citations

Alary, C. and H. Demougeot-Renard. 2010. Factorial kriging analysis as a tool for explaining the complex spatial distribution of metals in sediments. *Environmental Science Technology*. 44: 593-599.

Biswas, B. and B. Si. 2009. Spatial relationship between soil hydraulic and soil physical properties in a farm field. *Canadian Journal of Soil Science*. 89: 473-488.

Bocchi, S., A. Castrignanò, F. Fornaro, and T. Maggiore. 2000. Application of factorial kriging for mapping soil variation at field scale. *European Journal of Agronomy*. 13: 295-308.

Botros, F., T. Harter, Y. Onsoy, A. Tuli, and J. Hopmans. 2009. Spatial Variability of Hydraulic Properties and Sediment Characteristics in a Deep Alluvial Unsaturated Zone. *Vadose Zone Journal*. 8(2):276-289.

Chilés, J. P. and P. Delfiner. 2012. *Geostatistics: Modeling Spatial Uncertainty 2nd ed*. Hoboken, NJ: John Wiley & Sons, Inc.

Ciollaro, G. and N. Romano. 1995. Spatial variability of the hydraulic properties of a volcanic soil. *Geoderma*. 65: 263-282.

Corwin, D. and S. Lesch. 2005. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*. 46: 11-43.

Dane, T. and G. Topp. eds. 1994. *SSSA Book Series: 5, Methods of Soil Analysis, Part 4 Physical Methods*. Madison, WI: Soil Science Society of America.

Deurer, M. and W. Duijnisveld. 2000. Spatial analysis of water characteristic functions in a sandy podzol under pine forest. *Water Resource Research*. 36(10): 2925-2935.

Dobermann, A., P. Goovaerts, and T. George. 1995. Sources of soil variation in an acid Ultisol of Philippines. *Geoderma*. 68: 173-191.

Doussan, C. and S. Ruy. 2009. Prediction of unsaturated soil hydraulic conductivity with electrical conductivity. *Water Resource Research*. 45: W10408, doi:10.1029/2008WR007309.

Gambolati, G., M. Putti, P. Teatini, and G. Gasparetto Stori. 2006. Subsidence due to peat oxidation and impact on drainage infrastructures in farmland catchment south of Venice Lagoon. *Environmental Geology*. 49: 814-820.

Goovaerts, P.. 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*. 43: 597-619.

Groovaerts, P..1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biol Fertil Soils*. 27: 315-334.

Greminger, P., Y. Sud, and D. Nielsen. 1985. Spatial Variability of Field-measured Soil-water Characteristics. *SSSAJ*. 49(5): 1075-1082.

Grossman, R. and T. Reinsch. 2002. Bulk density and linear extensibility. In Dane, J.H., and G.C. Topp. Eds. *Methods of soil analysis, Part. 4*. Madison, WI: Soil Science Society of America.

Goulard, M. and M. Voltz. 1992. Geostatistical interpolation of curves: a case study in soil science. *Geostatistics Tróia*. 2: 805-816.

Hammel, K., J. Gross, G. Wessolek, and K. Roth. 1999. Two-dimensional simulation of bromide transport in a heterogeneous field soil with transient unsaturated flow. *European Journal of Soil Science*. 50: 633-647.

Jaynes, D. and E. Tyler. 1984. Using soil physical properties to estimate hydraulic conductivity. *Soil Science*. 138: 298-305.

Li, Y., D. Chen, R.E. White, A. Zhu, and J. Zhang. 2006. Estimating soil hydraulic properties of Fengqiu Country soils in the North China Plain using pedo-transfer functions. *Geoderma*. 138: 261-271.

Mallants, D., M. Binayak, J. Diederik, and J. Feyen. 1996. Spatial variability of hydraulic properties in a multi-layered soil profile. *Soil Science*. 161(3): 167-181.

Mallants, D., P. Tseng, N. Toride, A. Timmerman, and J. Feyen. 1997. Evaluation of multimodal hydraulic functions in characterizing a heterogeneous field soil. *Journal of Hydrology*. 195: 172-199. (A)

Mallants, D., B. Mohanty, A. Vervoot, and J. Feyen. 1997. Spatial analysis of saturated hydraulic conductivity in a soil with macropores. *Soil Technology*. 10: 115-131. (B)

Mertens, J., D. Jacques, J. Vanderborght, and J. Feyen. 2002. Characterisation of the field-saturated hydraulic conductivity on a hillslope: in situ single ring pressure infiltrometer measurements. *Journal of Hydrology*. 263: 217-229.

Mualem, Y. 1976. A New Model for Predicting the Hydraulic Conductivity of Unsaturated Conductivity of Unsaturated Porous Media. *Water Resources Research*. 12(3): 513-522

Nanos, N. and J. Martín. 2012. Multiscale analysis of Heavy metal contents in soils: Spatial variability in Duero river basin (Spain). *Geoderma*. 189-190: 554-562.

Pachepsky, Y.A., W.J. Rawls, and H.S. Lin. 2006. Hydropedology and pedotransfer functions. *Geoderma* 131:308–316.

Puckett, W., J. Dane, B. Hajek. 1985. Physical and mineralogical data to determine soil hydraulic properties. *Soil Science Society of America Journal*. 49: 831-836.

Radcliffe, D. and J. Simunek. 2010. *Soil Physics with HYDRUS: Modeling and Applications*. Boca Raton, FL: Taylor and Francis Group.

Russo, D. and W. Jury. 1987. A Theoretical Study of the Estimation of Correlation Scale in Spatially Variable Fields 1. Stationary Fields. *Water Resource Research*. 23(7): 1257-1268.

Rouse, J., R. Haas, J. Schell, D. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS symposium*. NASA SP-351: 309–317.

Schaap, M. and W. Bouten. 1996. Modeling water retention curves of sandy soils using neural networks. *Water Resources Research*. 32(10): 3033-3040.

Schaap, M. and F. Leij. 2000. Improved Prediction of Unsaturated Hydraulic Conductivity with the Mualem-van Genuchten Model. *SSSAJ*. 64: 843-851.

Scudiero, E., R. Deiana, P. Teatini, G. Cassiani, and F. Morari. 2011. Constrained optimization of spatial sampling in salt contaminated coastal farmland using EMI and continuous simulated annealing. *Procedia Environmental Sciences*. 7: 234-239.

Scudiero, E., P. Teatini, D. Corwin, A. Berti, R. Deiana, F. Morari. 2013. Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and apparent electrical conductivity. *Computers and Electronics in Agriculture*. 99: 54-64.

Simunek, J., M. Sejna, H. Saito, M. Sakai, and M. Th. van Genuchten, The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.0x, *Hydrus Series 3,* Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA, 2008.

Sollitto, D., M. Romic, A. Castrignanò, D. Romic, H. Bakic. 2010. Assessing heavy metal contamination in soils of Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena*. 80: 182-194.

Ünlü, K., D. Nielsen, J. Biggar, and F. Morkoc. 1990. Statistical Parameters Characterizing the Spatial Variability of Selected Soil Hydraulic Properties. *SSSAJ*. 54: 1537-1547.

van Genuchten, M. 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soil. *Soil Science Society of America Journal*. 44(5): 892-898.

Vauclin, M., D. Elrick, J. Thony, G. Vachaud, Ph. Revol, and P. Ruelle. 1994. Hydraulic conductivity measurements of the spatial variability of a loamy soil. *Soil Technology*. 7: 181-195.

Vereecken, H., M. Weynants, M. Javaux, Y. Pachepsky, M.G. Schaap, and M.Th. van Genuchten. 2010. Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: A Review. *Vadose Zone Journal*. 9: 795-820.

Voltz, M. and M. Goulard. 1994. Spatial interpolation of soil moisture retention curves. *Geoderma*. 62: 109-123.

Wackernagel, H.. 2003. *Multivariate Geostatistics 3rd ed*. Berlin, Heidelberg, Germany: Springer-Verlag.

Walczak, R., E. Rovdan, and B. Witkowska-Walczak. 2002. Water retention characteristics of peat and sand mixtures. *International Agrophysics*. 16: 161-165.

Weynants, M., H. Vereecken, and M. Javaux. 2009. Revisiting Vereecken Pedo transfer functions: Introducing a Closed-Form Hydraulic Model. *Vadose Zone Journal*. 8:86-95.

Wierenga, P., R. Hills, and D. Hudson. 1991. The Las Cruces Trench Site: Characterization, Experimental Results, and One-Dimensional Flow Predictions. *Water Resource Research*. 27(10): 2695-2705.

Wösten, J., Ya. Pachepsky, and W. Rawls. 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*. 251: 123-150.

Chapter 3: Spatial Prediction of Hydraulic Zones from Soil Properties and Secondary Data Using Factorial Kriging Analysis[3,4]

---

[3] To be submitted to *Geoderma*

[4] Authors: James Bevington, Francesco Morari and George Vellidis

Abstract

The development of pedotransfer functions (PTF) is an important topic in current soil science research because there is a critical need for incorporation of vadose zone phenomena into large scale climate models. Currently, much work is being done to improve predictions and reduce the requirement of intensive and invasive soil sampling. Soil measurements are inherently spatially dependent and therefore application of geospatial statistics provides an avenue for estimating soil properties. The objectives of this study are to investigate the application of factorial kriging analysis (FKA) and fuzzy-c means clustering to predict zones of hydraulic parameters which are similar to each other but differ from those of other zones. The work was conducted in a 20.8ha field located in the Po River delta in Italy. This region is in known to have paleo-channel structures and highly heterogeneous soils. FKA was applied to hydraulic parameters in one data set and soil physical properties in another data set at 4 depths. The mapped principal components (PCs) were used in a fuzzy-c means algorithm to define zones of like properties. Zones for predictive variables were compared to hydraulic zones by overlaying the two and calculating the areal prediction accuracy and also by direct comparison of the hydraulic parameters and curves for each zone. Areal accuracy ranged from 46.66% to 92.41%. For cluster performance, zones were able to distinguish between θ_s(saturated water content), *n* (shape parameter) and α (inverse of air entry) while θ_r (residual water content) and K_s (saturated conductivity) were not statistically different between the groups. This is also verified in comparisons of the hydraulic curves at 1000 soil water tension points. Water retention curves (WRC) were found to be significantly different at all tensions, effective saturation curves (Se) differ for the majority of tensions (except at 28cm), but unsaturated hydraulic conductivity curves (UHC) did not differ.

List of Abbreviations: BD, bulk density; ECa, apparent electrical conductivity; FKA, factorial kriging analysis; K_s, saturated conductivity; *n*, shape parameter; NDVI, Normalized Difference Vegetation Index; NFI, Non Fuzziness Index; PCA, principal component analysis; PTF, pedotransfer function; UHC, unsaturated hydraulic conductivity curve; WRC, water content curve; α, inverse of air entry; θ_r, residual water content; θ_s, saturated water content

1. Introduction

Being able to predict hydraulic properties of soil for large areas ranging from the field scale to the regional scale is a critical need in the field of soil science. This ability would improve accuracy of crop growth models as well as watershed transport and large scale climate models by introducing information about the vadose zone which is the interface between the ground water, terrestrial vegetation, and the atmosphere- all major components of these types of models. Measuring unsaturated hydraulic properties is hindered by cost, complexity and the immense time involved. For this reason, development of vadose zone predictive models also known as pedotransfer functions (PTF), has been underway for many years. In a review of these functions, Vereecken et al. (2010) suggested that new types of information should be evaluated for their ability to increase accuracy of PTF.

PTF are inherently spatially dependent because the data used to develop them is spatially dependent. Therefore, if the data were collected at a certain scale, the PTF is only valid for predictions at that scale (Vereecken et al., 2010). Upscaling and downscaling of these functions is an active area of research which attempts to adapt or develop methods to apply PTF at different scales from which they were originally developed (Vereecken et al. 2007). This task requires knowledge of the spatial variability of the soil under study. One method of upscaling involves aggregating a region into a single representative point with averaged properties or effective parameters (Vereecken et al. 2007).

The most commonly used models to describe the water retention curve (WRC) and the unsaturated hydraulic conductivity curve (UHC) are the van Genuchten (1980) and Mualem (1976) models, respectively. Many PTF attempt to predict the parameters used in these models from soil physical properties such as texture, bulk density, chemical properties such as soil organic content (SOC), and others. Once the parameters are known, values of the water content or hydraulic conductivity can be calculated for any given tension. Doussan and Ruy (2009) developed a PTF to use directly measured

electrical conductivity (EC) as input to predict UHC. Directly measuring EC could prove difficult, especially for large scale studies. However, remotely sensed apparent EC (ECa) is currently used on large sites. Moreover, ECa has been used extensively in spatial studies of soil (Corwin and Lesch, 2005).

PTF are also only as good as the data used to develop them. Many authors have reported discrepancies between lab measured and field measured hydraulic properties and parameters (Wierenga et al., 1991; Mallants et al., 1997b). In fact, Mertens et al. (2002) found that the main source of variation in measured hydraulic parameters was due to measurement uncertainty, not variability in the sample population. The leading cause is likely heterogeneity of the soil. Mallants et al., (1997b) proved this to be true for measurements of K_s. Because this heterogeneity is also likely to be spatially dependent, spatial statistics can be used to account for these discrepancies.

Studies have reported on the spatial relationships and scale dependence between hydraulic parameters and soil physical properties (Biswas and Si, 2009; Ciollaro and Romano 1995; Mallants et al., 1997a; Hammel et al., 1999; Russo and Jury, 1987; Deurer et al., 2000; Botros et al. 2009). Though geospatial statistics have been used in scaling of PTF, there are few, if any, attempts to use spatial statistics to develop PTF (Vereecken et al. 2010). Factorial kriging analysis (FKA) is a multivariate geospatial analysis that is useful for identifying underlying processes which cause or affect variability in the variable space. This is accomplished in much the same way that principal component analysis (PCA) reduces a variable space into a few factors which explain the entire set. FKA is essentially a spatial PCA but it also separates relationships based on scale. In brief, FKA consist of three steps, defining variogram and cross-variogram models which meet the constraints of the linear model of coregionalization (LMC), regionalized PCA, then mapping of regionalized PCs. For a detailed explanation of the theory behind FKA see Goovaerts, (1992), Chilés and Delfiner, (2012) and Wackernagel (2003).

By using FKA, studies have identified sources of variation and chemical leaching in soil (Dobermann et al., 1995), distinguished between anthropogenic and natural sources of trace elements in soil (Sollitto et al., 2010; Nanos and Martin, 2012), identified the spatial effects of an earthquake on aquifer properties (Lin et al., 2004), among many other uses. With regards to hydraulic properties, Bocchi et al. (2000) used FKA to study relationships between water content at three tensions and some soil physical properties. They also identified manure spreading as a source of variability. Biswas and Si (2009) is the only known application of FKA to study relationships between van Genuchten hydraulic parameters and soil properties.

The goal of this study is to investigate the use of FKA to make spatial predictions about hydraulic parameters from soil physical and remotely sensed or exhaustively known datasets by defining and predicting zones of hydraulic parameters. The zones should group curves which are similar to each other but differ from those of other zones.

2. Methods

2.1 Study Site description

The study site is located on the southern edge of the Venice lagoon in the Po River delta. This is a section of land which was reclaimed between 1892 and 1967 (Gambolati et al., 2006). Reclaiming of land from the lagoon requires construction of an earth retaining wall and collection and removal of water to maintain a dry vadose zone. Water collection is achieved through a system of open ditches which route water to a pumping station where it is removed. The soil is highly variable in terms of texture and bulk density and there are 2 well preserved paleo-channels with high sand content which cross the field. On the northern edge of the 26.5 ha field, there are three constructed river canals including the Morto Canal. When the Morto Canal was constructed, mineral soil of alpine origin was

spread over the soil of lagoon origin on the northern portion. Soil on the northern edge is classified as Fluvaquentic Endoaquepts fine-silty mixed mixed, calcareous, mesic while soil on the southern edge is classified as Typic Sulfisaprists euic, mesic by the USDA (1998) system. The field is used for agricultural purposes including annual tillage to a depth of approximately 30 cm and cultivation of *Zea mais* L. Due to high precipitation, core samples (see below) were only taken in the 20.8 ha section indicated in figure 3.1 with a red outline. Figure 3.1 also depicts the features described above.

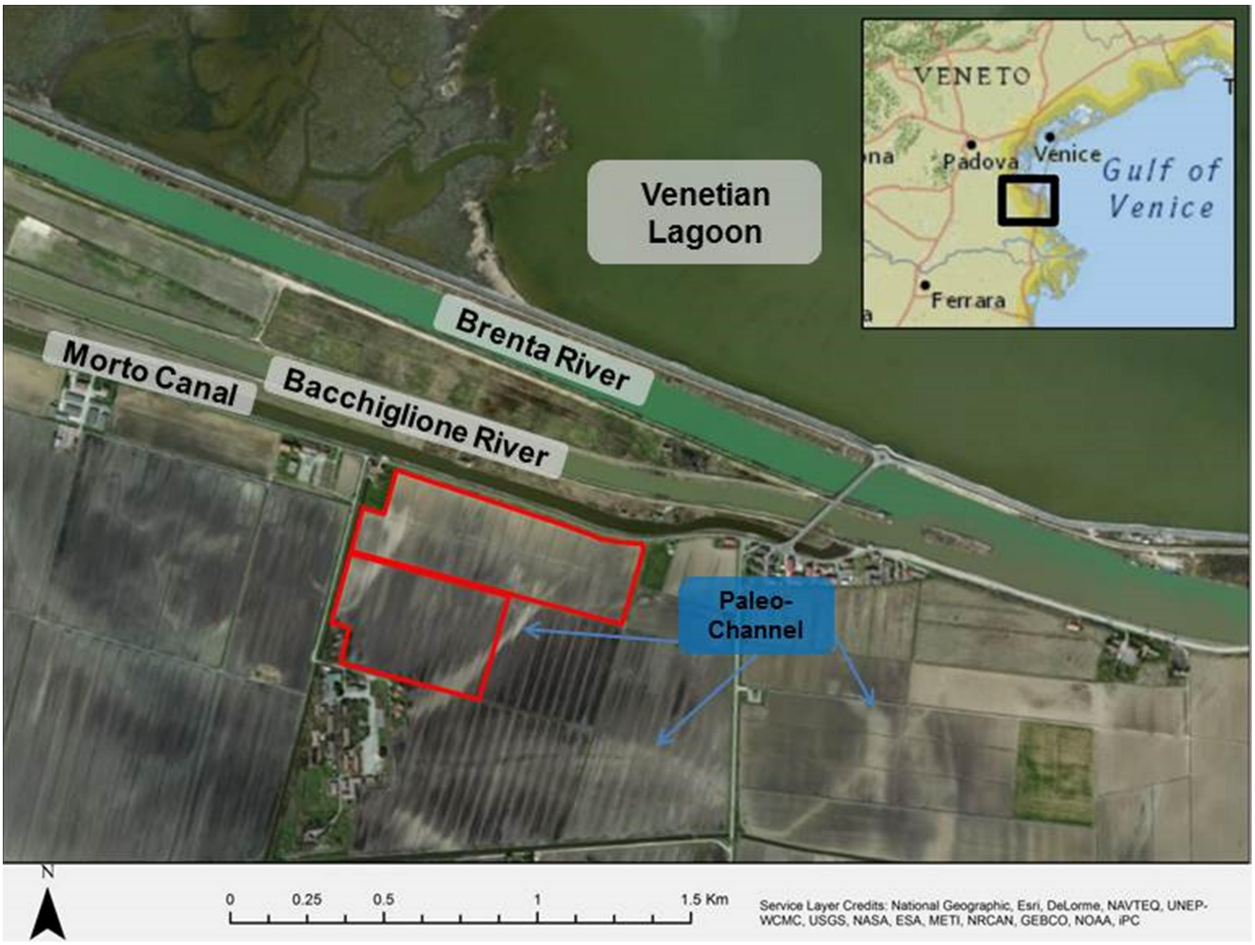


Figure 3.1. Map depicting the 20.8 ha study area in red, 3 river canals, Venetian Lagoon, and paleo-channels marked in blue. The site is located at the southern edge of the Venitian Laggon where paleo-channels are known to exist.

2.2 Exhaustive Data

Four sets of exhaustive data were collected to be used as secondary data for the geospatial analysis. The use of secondary data can improve prediction results of primary variables (Wackernagel, 2003). The 4 sets are apparent electrical conductivity (ECa) at 0-0.75 m (ECa075) ECa at 0-1.5 m (ECa150), elevation, and bare soil reflectance as measured with the Normalized Difference Vegetation Index (NDVI). These were collected on the entire 26.5 ha of the site. ECa was measured separately for each depth using a CMD-1 frequency-domain electric induction sensor (GF Instruments, Brno, Czech Republic) in different configurations for the different depths. The CMD-1 collects 2 measurements per location while moving and more if the instrument is stopped. The multiple measurements for each location were averaged together. There were 18053 measurements at 9017 locations for ECa075 and 20471 measurements at 10221 locations for ECa150. Elevation measurements were made at 1564 locations using a Trimble FM 1000 CNH GPS receiver with real time kinematic (RTK) differential correction (Trimble Navigation Ltd., Sunnyvale, CA, USA). Bare soil NDVI was calculated from measurements of 590 nm (VIS) and 880 nm (NIR) wavelength reflectance using $NDVI = (NIR - VIS)/(NIR + VIS)$ (Rouse et al., 1973). The reflectance measurements were made using an APS1-CropCircle (Holland Scientific, Lincoln, NE, USA) at 10214 locations. Using ISATIS v13.01 (Geovariances and Ecole des Mines de Paris, Avon Cedex, France), each dataset was interpolated to a 1m resolution grid of 750 by 550 nodes. The interpolation was done by transforming the raw data, de-trending, then kriging the residuals (ordinary kriging). Grid values were exported to Matlab (MathWorks, Natick, MA) where a PCA was used to reduce the 4 variables into 1. This new variable is the first principal component (F1) and was used as a secondary variable.

2.3 Soil Sampling

Scudiero et al. (2011) described the sampling location selection process which utilized a simulated annealing technique from an ECa survey to capture the maximum variance. Disturbed

samples were taken at 123 locations at depths of 0-0.15 m, 0.15-0.45 m, and 0.45-0.80 m. At, 50 of these locations, undisturbed samples were also taken at depths of 0.05-0.11 m, 0.25-0.31 m, 0.45-0.51 m, and 0.65-0.71 m. Texture was measured on disturbed samples with a Mastersizer 2000 (Malvern Instruments Ltd, Great Malvern, UK). The undisturbed samples were used to measure bulk density using the Core method (Grossman, R. and T. Reinsch, 2002)and hydraulic measurements using the Wind method (Dane and Topp, 1994) with a Ku-pF Apparatus DT 04-01 (Umwelt-Geräte-Technik GmbH, Müncheberg, Germany). The Apparatus uses tensiometers with a working range of 0-90 kPa and measurements were taken every 10min. Evaporation was stopped when sensor limits were reached, or when measurements were unreliable. Van Genuchten (1980) and Mualem (1976) parameters were fit to the data using an inverse method in HYDRUS 1D v4.12 (Simunek et al., 2008), a program which is commonly used for 1D transport modeling in unsaturated soil. Before inversion, measures hydraulic data were binned and averaged on a $\log_2$ scale which reduced the number of data points while maintaining representation of low tension points. Stopping criteria of 0.001% for water content tolerance and 1cm head for pressure head tolerance was used.

2.4 Spatial Statistics

Though a normal distribution is not required for kriging, highly skewed distributions can produce poor results (Alary and Demougeot-Renard, 2010). Hydraulic parameters have been found by many to have extreme, very non-normal distributions (Mallants et al., 1997b; Mallants et al. 1996; Bostros et al. 2009). Therefore, a Hermitian transformation algorithm was applied to all variables (Wackernagel, 2003). A first or second order polynomial was fit to each transformed variable to ensure that no global trends were present in the data. ISATIS was used for most of the geostatistical work but some tools from ArcMap 10.1 (ESRI, Redlands, CA, USA) were also used.

Experimental variograms and cross-variograms were computed for hydraulic parameters and predictor variables separately. The only common variable between the groups is the secondary variable. Ideally, there would be no common variables. However, maps made without the use of a secondary variable were poor and not representative; therefore the secondary variable was retained. In order to isolate the effects of the secondary variables on prediction, an analysis was conducted using just the secondary variables. Variogram models were fit to experimental data with a gradient method optimization algorithm in ISATIS (ISTATIS Help File, Geovariances and Ecole des Mines de Paris, 2013). Sum of squared residuals, Akaike criterion and Bayesian Information criterion were used to determine the fit quality of the models. All models were constrained by the linear model of coregionalization (LMC) which requires all variograms and cross-variograms in a model to be fit with the same set of ranges (Wackernagel, 2003). For each model, up to three structures were considered, nugget plus two more.

Once models were fitted, factorial kriging analysis (FKA) was conducted. FKA is a regionalized PCA technique which is useful for identifying spatially dependent relationships between variables and mapping factors which represent the underlying processes which cause the spatial distributions. This technique has been widely described in by Goovaerts, (1992), Chilés and Delfiner, (2012) and Wackernagel (2003). There are three main steps in FKA, selecting a variogram and cross-variogram model, PCA of the regionalized sill values, and mapping of the regionalized factors. Factors were mapped to a 1m resolution grid using a multi-collocated cokriging method which improves map representation by utilizing information from secondary data. Multi-collocated cokriging has greatly reduced computational requirements than other cokriging schemes which enables easier use of large secondary datasets (Wackernagel, 2003).

2.5 Clustering and Prediction Evaluation

The mapped regionalized factors (grid node values) inside of the study were exported to Matlab for cluster analysis. The first factor for each structure or the first two factors if there was only one structure (excluding nugget) were used as input data. Cluster analysis has been used by many to define management zones from continuous variables. Even though this type of analysis does not include spatial portions of the data, it is still effective because values tend to be similar to those nearby and continuous zones can be obtained. A fuzzy c-means algorithm was used to determine clusters (Bezdek 1981). A value of 1.3 was used for the fuzziness partition weighting exponent and convergence criteria of 1e-5 or 200 iterations. These values were chosen based on work by Odeh et al. (1992) and Fridgen et al. (2004). For this method, the number of clusters is given at the beginning then an indication of how well each point fits into each group is calculated. However, the appropriate number of clusters is not known at the beginning. As suggested by Roubens (1982), clustering was calculated for 2-6 clusters then performance was evaluated by the Non Fuzziness Index (NFI) $NFI = (c(\sum\sum U^2/n) - 1)/(c - 1)$ and Entrolpy function $H = 1 - (-\sum\sum U \log_a U\ /n)$ where *c* is the number of clusters, *U* is the fuzziness membership function, and *n* is the number of data points (Roubens, 1982). For both indexes, a value of 1 indicates perfect clusters and 0 indicates no clusters. These indexes were summed and the number of clusters which produced the maximum sum was selected.

Prediction was evaluated in two ways. First, the zones for the prediction data were laid over the zones for the hydraulic data. The percentage of nodes whose membership was correctly predicted was calculated. Second, hydraulic parameters and hydraulic curves were sorted into their predicted clusters and then comparison between the predicted zones was conducted. A one-way ANOVA of the hydraulic parameters was calculated followed by Tukey's honestly significant difference (HSD) criterion to determine significant difference at a probability of 0.05. The one-way ANOVA assumes that the response variables have a normal distribution and groups have equivalent variances. While the one-way ANOVA is

relatively insensitive to deviations from normality (as long as the population is normal) it is sensitive to heteroscedasticity when comparing uneven sample sizes (Glass et al., 1972). To ensure reliable results, transformed hydraulic parameters were used in the one-way ANOVA and homoscedasticity was verified at P=0.05 with a Levene's test. WRC, UHC, and effective saturation curve were calculated from the hydraulic parameters at 1000 tensions ranging from 0 kpa to 800 kpa on a $\log_2$ scale. Then, curves were compared at each of these tension points. The distribution of each cluster was checked for normality using a one-sample Kolmogorov-Smirnov test at P=0.05 and a Levene's test was used to check for homoscedasticity at P=0.01. Then, a one-way ANOVA and Tukey's HSD criterion were used to determine statistical significance at P=0.05.

3. Results and Discussion

3.1 Factorial Kriging Analysis

3.1.1 Variography

For the hydraulic data, θ_s exhibited strong spatial structure. α, n, θ_r, and K_s, had some structure but also, a lot of nugget and they were sensitive to changes in bin size. l had very poor spatial structure so it was dropped from analysis. For the prediction variables, sand, clay, and bulk density had strong to moderate spatial structure. The texture variables were very stable because they were represented at more locations. There is not enough space to include variogram models but a summary of the models is given in table 3.1. All depths except 28 cm had two spherical structures. 28 cm had only one spherical structure. Short scale ranges tended to be around 87 m for the hydraulic models and 93 m for the predictive models. Long scale ranges for the hydraulic models were around 218 m and 231 m for the predictive models. The similarity in the ranges between the hydraulic and predictive models indicates that these variables are likely linked to the same underlying processes.

Table 3.1. Hydraulic and Predictive Model Ranges

	Hydraulic		Predictive	
Depth (Nom.)	**Short[1] (m)**	**Long[1] (m)**	**Short[1] (m)**	**Long[1] (m)**
8	102.0	207.8	96.6	256.6
28		200.5		232.9
48	77.0	234.5	74.1	193.0
68	82.1	227.3	109.5	242.2

1) Short and Long are the short and long scale ranges for the spatial models.

3.1.2 Regionalized PCA

Results of the regionalized PCA are presented in tables 3.2A-H. Many of the principal components (PCs) had Eigenvalues less than one. Typically, these would be dropped from analysis because they do not hold much information for describing the entire variable space. However, in this case, they were retained because they do hold information about the relationship structure between variables. Also, hydraulic parameters and soil physical properties are known to be highly variable in space. In addition, this is a highly variable soil. Capturing this variability fully would require an impractical amount of samples. Instead, criteria on PC retention were relaxed as to retain information that would likely be obvious if more samples were measured. It should also be noted that even though the Eigenvalues are low, these PCs still account for the majority of the variance at their particular scales. With the exception of the second PC at long range (PC2L) at 28 cm and PC1L at 48 cm, the PCs appear to represent the hydraulic parameters more so than the secondary variables. This is good because the goal is to predict hydraulic parameters, not secondary variables. The hydraulic models are characterized by loadings of θ_s at short scale and *n* at long scale. The predictive models are dominated by bulk density at short scale and texture at long scale.

Table 3.2A. 8 cm Hydraulic Model Regionalized PCs

	K_s	α	*n*	θr	θs	Elev	Eigen	% Var
PC1S	-0.199	0.798	-0.168	0.251	0.461	-0.145	0.731	85.2
PC1L	0.048	-0.044	-0.177	0.501	0.737	-0.413	0.703	85.2

Table 3.2B. 8 cm Predictive Model Regionalized PCs

	BD	Clay	Sand	Elev	Eigen	% Var
PC1S	0.843	-0.308	0.417	0.142	0.707	90.2
PC1L	0.334	0.652	-0.680	-0.002	0.654	63.3

Table 3.2C. 28 cm Hydraulic Model Regionalized PCs

	K_s	α	*n*	θr	θs	F1	Eigen	% Var
PC1L	0.254	0.328	-0.262	0.180	0.508	0.685	2.623	81.5
PC2L	0.569	0.168	-0.552	-0.306	0.094	-0.492	0.542	16.8

Table 3.2D. 28 cm Predictive Model Regionalized PCs

	BD	Clay	Sand	F1	Eigen	% Var
PC1L	0.542	-0.189	0.153	-0.805	2.213	80.4
PC2L	-0.107	0.606	-0.348	-0.707	0.534	19.4

Table 3.2E. 48 cm Hydraulic Model Regionalized PCs

	K_s	α	*n*	θr	θs	F1	Eigen	% Var
PC1S	-0.549	-0.520	-0.092	-0.461	-0.379	-0.2536	2.022	80.0
PC1L	-0.026	0.212	-0.569	-0.025	0.250	0.7539	2.142	82.3

Table 3.2F. 48 cm Predictive Model Regionalized PCs

	BD	Clay	Sand	F1	Eigen	% Var
PC1S	0.891	-0.104	0.176	-0.406	1.090	94.4
PC1L	-0.175	0.561	-0.596	0.547	2.127	83.2

Table 3.2G. 68 cm Hydraulic Model Regionalized PCs

	K_s	α	*n*	θr	θs	F1	Eigen	% Var
PC1S	0.009	0.328	0.153	0.925	-0.020	0.112	0.854	58.9
PC1L	0.201	0.505	-0.509	0.055	0.469	0.471	2.838	70.5

Table 3.2H. 68 cm Predictive Model Regionalized PCs

	BD	Clay	Sand	F1	Eigen	% Var
PC1S	0.638	-0.269	0.447	-0.567	0.946	99.9
PC1L	0.313	-0.556	0.531	-0.558	2.681	78.8

3.1.3 Spatial Distribution of PCs

Overall, maps of the hydraulic parameter PCs resembled bulk density maps at short scale and the paleo-channels at long scale. The predictive PC maps also resembled bulk density at short scale and the paleo-channels at long scale. As an example hydraulic and predictive maps for 8 cm are presented in figures 3.2A-F. The predictive maps have a more defined spatial distribution and have obviously benefited from the higher density of texture points. However, the distributions are similar which is a second indication that these variables are related through the same underlying processes.

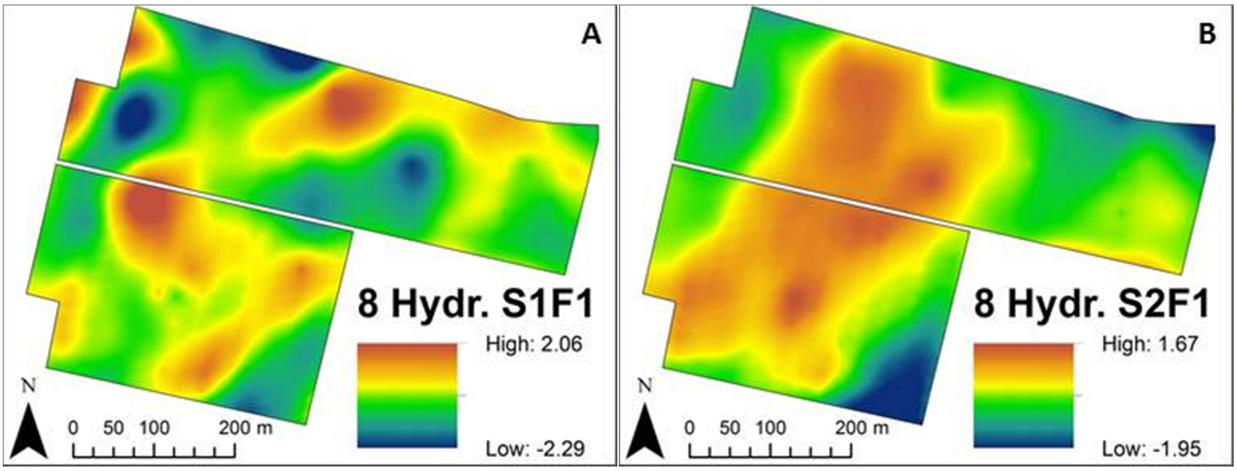


Figure 3.2A-B. Maps of short range (S1F1) and Long range (S2F1) retained PCs from the hydraulic model.

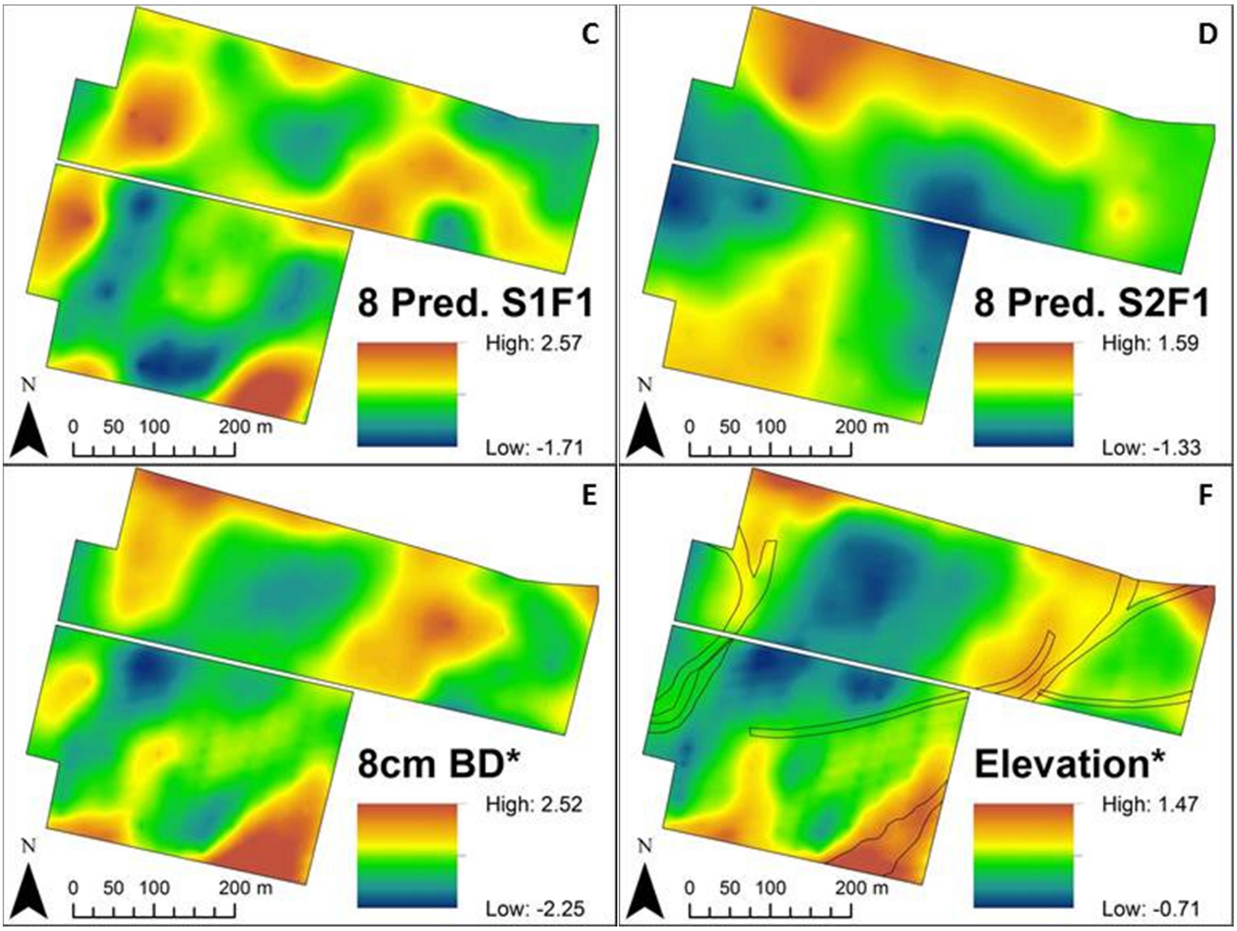


Figure 3.2C-F. C-D) Maps of retained PCs from the predictive model. E) Bulk density. F) Well preserved paleo-channels (outlines) and elevation. Maps in the left column have similar distributions and maps in the right column have similar distributions though there are some deviations.

3.2 Zone Predictions

3.2.1 Zone Definition and Spatial Evaluation

For each depth, the number of clusters was the same for the hydraulic model and the predictive model. Except for 28 cm, the optimal number of clusters was 3. For 28 cm, the optimal number was 2. After the optimal number of clusters was selected, prediction zones were laid over hydraulic zones and the number of grid nodes accurately predicted was calculated. The zones are shown in figures 3.3A-H. The highest accuracy was at 28 cm at least in part because it only had 2 zones. Of the other depths, 48 cm was the best at 84.14%. This may just reflect that PC 2L at 28 cm and PC 1L at 48 cm and represent the secondary variable which is also included in the predictive model. Regardless, the lowest prediction,

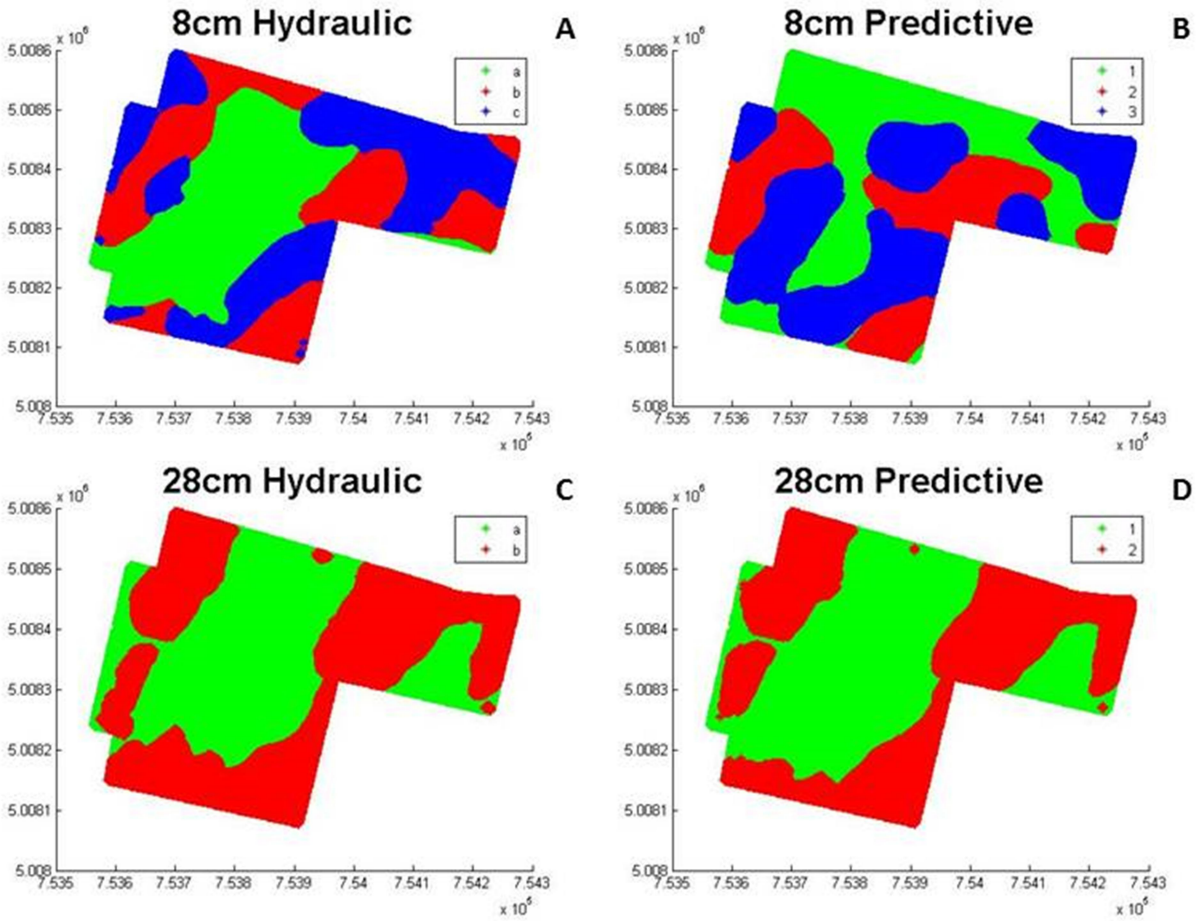
46.66% was still a 39.81% increase over a random prediction. The predictions also have another major advantage over a random prediction in that they are spatially clustered as well as clustered by attribute. For spatial modeling this is a key point. Table 3.3 is a summary of the cluster selection analysis.

Table 3.3. Summary of cluster analysis using all prediction data

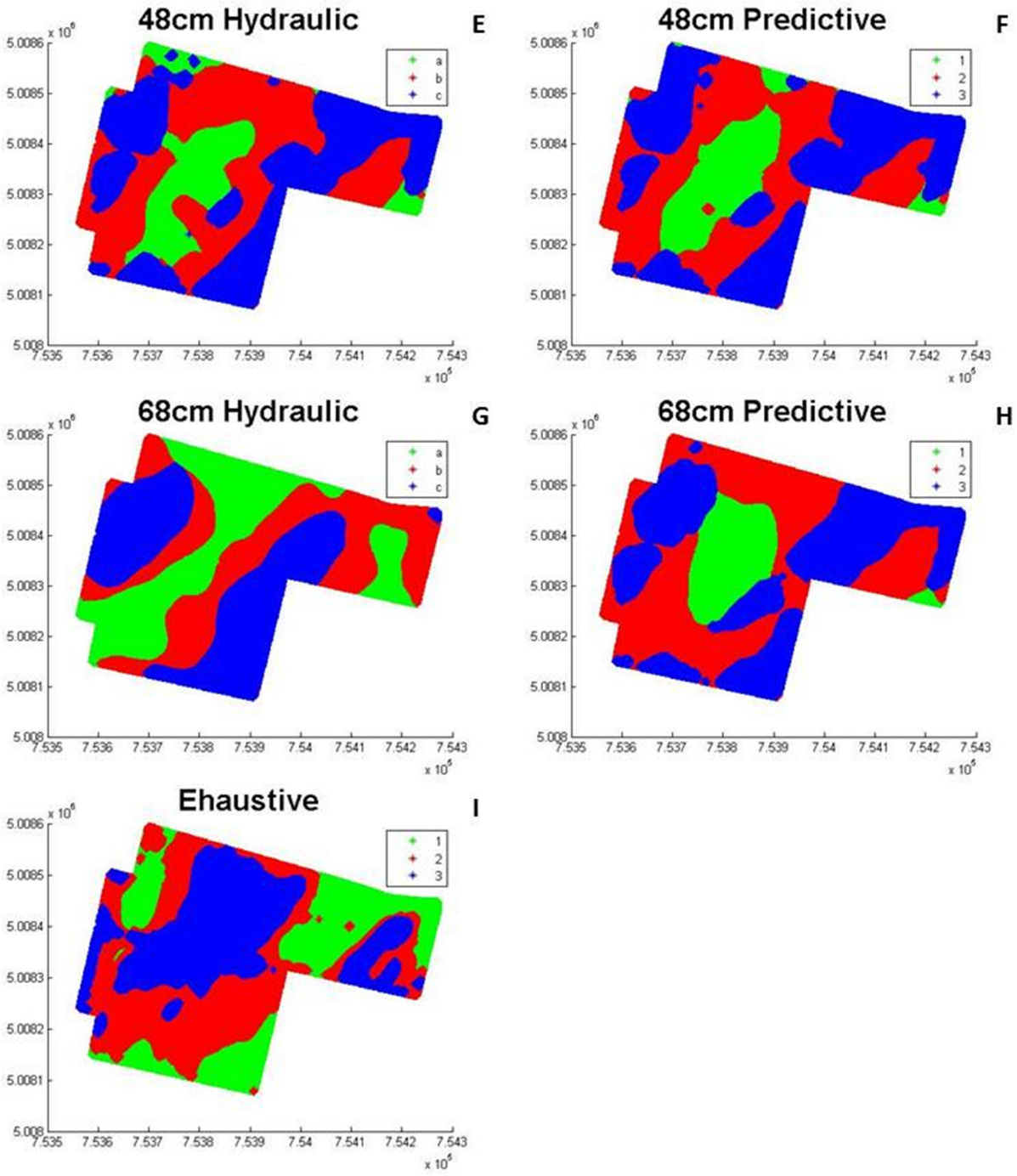
	Hydraulic Clusters	Prediction Clusters	% Prediction[1]	% inc. over random[2]
8cm	3	3	52.16	56.50
28cm	2	2	92.41	84.82
48cm	3	3	84.14	152.45
68cm	3	3	46.66	39.81

1) % Prediction is the portion of grid nodes that were accurately predicted.
2) % inc. over random = (% Prediction-(100/Prediction Clusters))/(100/Prediction Clusters)

In an attempt to understand the influence of the secondary variables in the above predictions, a similar analysis was conducted using only the exhaustive data. All 4 variables, ECa075, ECa150, Elevation, and NDVI were clustered using the fuzzy-c means algorithm and laid over the hydraulic zones. The cluster number was set to the same number of clusters as the hydraulic and predictive models even though the optimal number of clusters for the exhaustive model was 2. At 8cm, 28cm, 48,cm, and 68cm the number of nodes accurately predicted was 52.84%, 93.35%, 59.36%, and 46.22% respectively. Except for 48cm, this was roughly the same as the predictive model. So, either the secondary variables dominate the analysis or they represent the same underlying processes that are responsible for the variance in the hydraulic properties and soil physical properties, or a combination of these. Ideally, the secondary variables represent the primary variables well because this is their purpose in cokriging. However, it cannot be ruled out that they have an overpowering influence in the analysis and all that is revealed is that variable X predicts variable X because they are included in both models and the prediction error is the noise introduced by including other variables or error from interpolation.



Figures 3.3A-D. A. C.) Maps of zones for hydraulic parameters. B. D.) Maps of zones for predictive variables. Side by side comparisons reveal many similar structures between the two sets of models though there are some deviations. Axes are UTM coordinates.



Figures 3.3E-I. E. G.) Maps of zones for hydraulic parameters. F. H.) Maps of zones for predictive variables. Side by side comparisons reveal many similar structures between the two sets of models though there are some deviations. I.) Map of zones predicted by 4 secondary variables. This map has zones with many of the same features as the others. However, it has many differences also. Axes are UTM coordinates.

3.2.2 Attribute Evaluation

The above analysis illustrates some important points, but, it does not allow for strong conclusions to be drawn – mostly because the same secondary variables were included in both the predictor and predicted data groups. A stronger analysis, would distinguish the effects of the secondary variable from the effects of prediction error. Spatially, there are not enough data, especially at short lags, to create accurate zones for the highly variable hydraulic group. However, this was accomplished by investigating the hydraulic properties directly. Hydraulic parameters and hydraulic curves were separated into their predicted groups, and then compared to the other groups. Results for the hydraulic parameter comparison are in table 3.4 and results for the curve comparison are in table 3.5. Plots of these curves can also be seen in figures3. 4A-L.

All parameters had equivalent variances at P=0.01 but at P=0.05, θ_r at 28 cm and 68 cm and α at 8cm rejected the hypothesis of homoscedasticity. Overall, the groups did well at distinguishing hydraulic parameters except for K_s and θ_r. These two parameters tended to be difficult to model in the variograms. K_s is known to have a autocorrelation range much shorter than what this sampling scheme could capture. Also, inversion results often returned large confidence limits on K_s in comparison to the other parameters used. This uncertainty may account for the large nugget effect on K_s which leads to its poor spatial structure. θ_r is considered by many to be a fitting parameter and little physical significance (Radcliffe and Simunek, 2010). Because of this, it is unlikely that a strong spatial structure would be present. The groups did best at distinguishing θ_s and n. θ_s had the strongest differences while n had the most significant differences. The zones were able to identify one differing group for α for all depths except 68 cm.

At all 1000 tension points on the three hydraulic curves, each group was tested for normality using the Kolmogorov-Smirnov test (P=0.05) and equivalent variances using Levene's test (P=0.01). At all

depths, curves, and points the null hypothesis of normality was accepted. For the most part, the null hypothesis of homoscedasticity was also accepted; the major exclusion being the bottom half of the Se curve at 8 cm. A one-way ANOVA followed by a multiple comparison test was conducted at each tension point. The number of points where at least one group was significantly different from another group was tallied. The results are shown in table 3.5. It is important to note that this comparison did not consider the parameter confidence limits for each point. At all depths the WRC had at least one significantly different group at all tension points. This is confirmed in figures 3.4E-H; the WRC for each group are very different mostly by translation from differences in θ_s. The Se curves, except at 28 cm, also had many tension points with a significantly different group. For the Se curves, it is impossible to have all 1000 points be significantly different because the curves are fixed to the same values at the lower and upper boundaries. While the groups were good at distinguishing WRC and Se they performed poorly at distinguishing UHC curves. The best UHC separation occurred at 8cm with only 27.2% of the tension points having a significantly different group. Figures 3.4I-L illustrate this, as the UHC are almost identical with the only separation occurring around the air entry point. This is likely due to a poor distinction in K_s.

Table 3.4A. 8 cm Comparison of hydraulic parameters between predicted groups

8cm	P>V[1]	P>F[2]	1x2[3]	1x3[3]	2x3[3]
K_s	0.6494	**0.0101**			X
α	**0.0194**	**0.0020**		X	
n	0.1396	**0.0044**	X		
θr	0.1033	0.4663			
θs	0.2928	**5.36E-05**		X	X

Table 3.4B. 28 cm Comparison of hydraulic parameters between predicted groups

28cm	P>V[1]	P>F[2]	1x2[3]
K_s	0.5268	0.0883	
α	0.6591	**0.0068**	X
n	0.9241	**0.0220**	X
θr	**0.0458**	0.6259	
θs	0.1474	**0.0006**	X

Table 3.4C. 48 cm Comparison of hydraulic parameters between predicted groups

48cm	P>V[1]	P>F[2]	1x2[3]	1x3[3]	2x3[3]
K_s	0.6767	**0.0343**	X		
α	0.3491	**0.0111**	X		
n	0.0761	**0.0015**		X	X
θr	0.4444	**0.0038**	X		
θs	0.6750	**0.0003**	X		

Table 3.4D. 68 cm Comparison of hydraulic parameters between predicted groups

68cm	P>V[1]	P>F[2]	1x2[3]	1x3[3]	2x3[3]
K_s	0.3615	0.3043			
α	0.0841	**0.0792**			
n	0.6263	**0.0042**		X	X
θr	**0.0012**	**0.0441**			X
θs	0.2294	**0.0002**		X	

1) P>V is the probability of accepting the null hypothesis of equivalent variances. All parameters accept the null hypothesis at P=0.01.
2) P>F is the probability of accepting the null hypothesis of the F-test of ANOVA. Bold font denotes a rejected null hypothesis at P=0.05.
3) The columns to the right represent ad hoc comparisons between each group. An X denotes a significant difference at P=0.05.

Table 3.5A. 8 cm Comparison of Se, WRC, and UHC curves at 1000 tensions

8cm	% Normal[1] (P=0.05)	% Equal Var[2] (P=0.01)	% Different[3] (P=0.05)
Se	100	50	86.8
WRC	100	100	100
UHC	100	99.8	27.2

Table 3.5B. 28 cm Comparison of Se, WRC, and UHC curves at 1000 tensions

28cm	% Normal[1] (P=0.05)	% Equal Var[2] (P=0.01)	% Different[3] (P=0.05)
Se	100	100	1.2
WRC	100	100	100
UHC	100	100	19.1

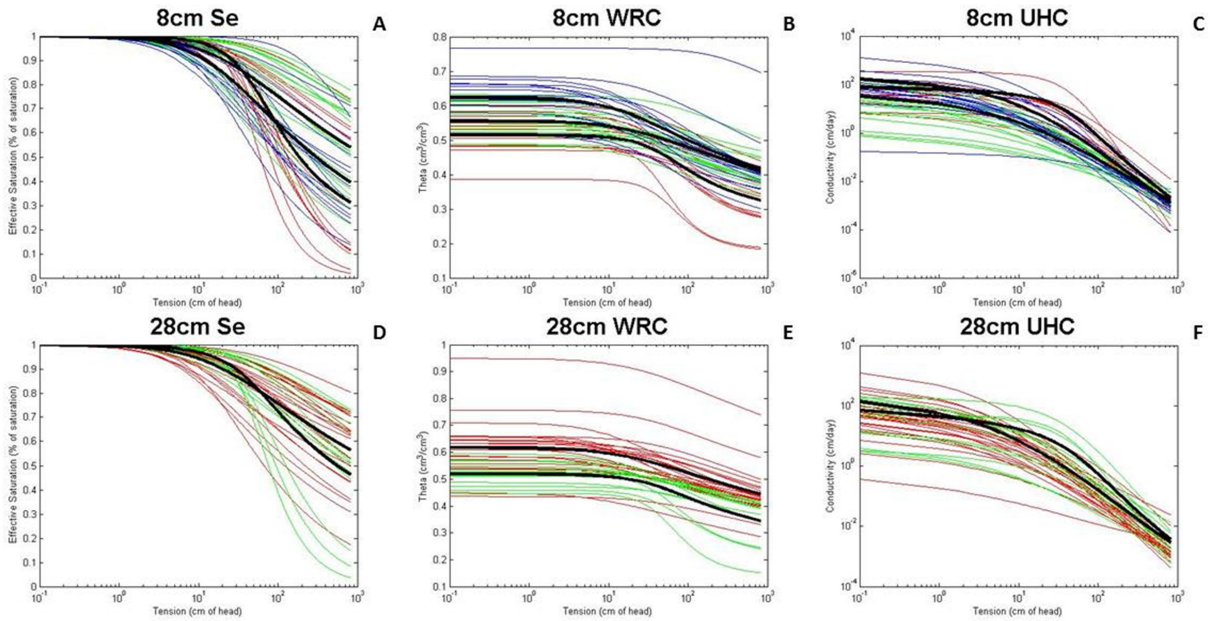
Table 3.5C. 48 cm Comparison of Se, WRC, and UHC curves at 1000 tensions

48cm	% Normal[1] (P=0.05)	% Equal Var[2] (P=0.01)	% Different[3] (P=0.05)
Se	100	99.7	99.5
WRC	100	100	100
UHC	100	100	0.6

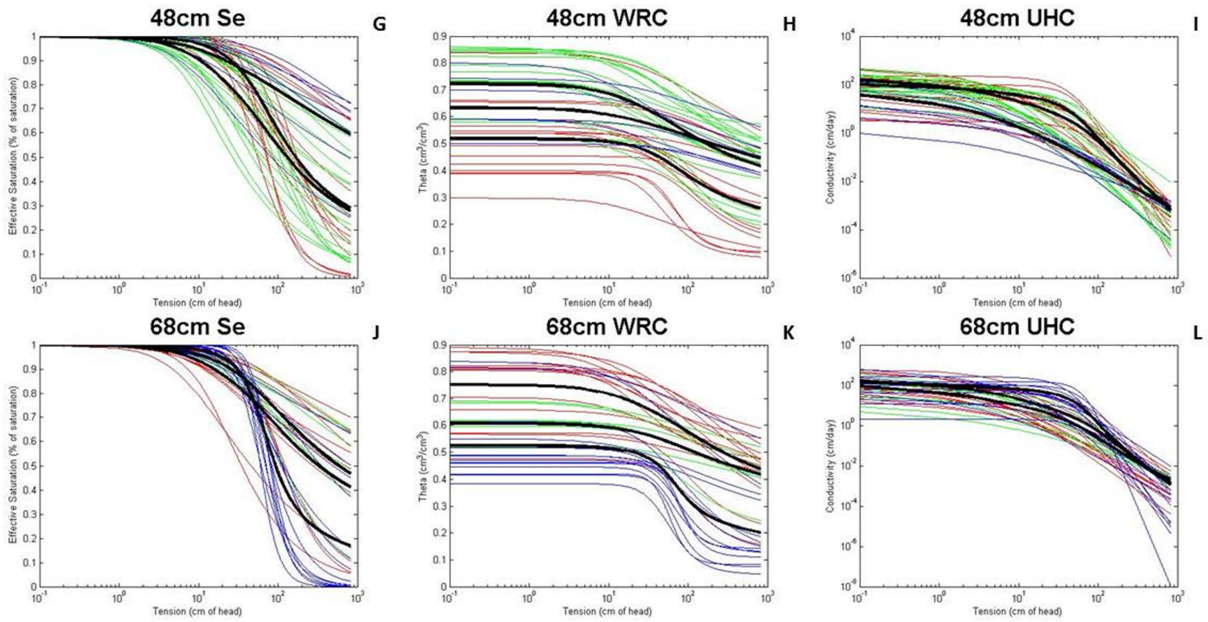
Table 3.5D. 68 cm Comparison of Se, WRC, and UHC curves at 1000 tensions

68cm	% Normal[1] (P=0.05)	% Equal Var[2] (P=0.01)	% Different[3] (P=0.05)
Se	100	100	94.2
WRC	100	100	100
UHC	100	100	5.9

1) % Normal represents the portion of points that accepted the null hypothesis of normality.
2) % Equal Var is the portion of points that accepted the null hypothesis of equivalent variances.
3) % Difference is the portion of points that had at least one significantly different group.



Figures 3.4A-F. Plots of hydraulic curves grouped by prediction zone. Colors represent the different zones as shown in figure 3.3. Black lines represent the averaged curve for each zone. Each colored line is a unique lab measured sample. Colors do not match zone colors from figures 3.3 (red curves are not necessarily from red zones). It is clear from visual inspection which curves are well distinguished and which are not.



Figures 3.4G-L. Plots of hydraulic curves grouped by prediction zone. Colors represent the different zones as shown in figure 3.3. Black lines represent the averaged curve for each zone. Each colored line is a unique lab measured sample. Colors do not match zone colors from figures 3.3 (red curves are not necessarily from red zones). It is clear from visual inspection which curves are well distinguished and which are not.

4. Discussion and Conclusions

The method presented here worked well at distinguishing WRC and moderately at distinguishing Se curves. UHC were not distinguished well by this method. This may be because K_s is not well represented in the spatial models. The predicted zones were an improvement over a random attribute selection but more importantly, they were spatially contiguous. For upscaling purposes and for use with large models defining zones may be more useful than having a continuous prediction because of the computational power required (Pietro Teatini, personal communication).

It is important to note that just because the algorithm produces zones, does not mean that these are the best, most representative zones (Roubens, 1982). In this study, the predicted zones did produce significantly different groups. However, it is highly likely that different zones might represent differences in the hydraulic data better than the ones presented here. Perhaps a comparison between

the predicted clusters and clusters defined directly on the hydraulic parameters excluding spatial characteristics could serve as an indication of the best achievable clusters.

Zones are useful from a modeling and spatial upscaling view because they spatially aggregate properties. If for nothing else, the approach proposed in this study could be used to direct soil sampling for hydraulic measurements. In theory, one sample per zone could be used as a representative curve for that zone. However, if a more detailed description is needed, many samples could be taken then, scaled using any hydraulic scaling method such as functional normalization, Miller-Miller, and Leverett (Vereecken et al., 2007). This option adds flexibility to the approach and allows for a tailored application to the study needs.

One very positive outcome of this approach is the strong identification of θ_S groups. θ_S is strongly connected to short scale processes. In order to accurately represent these processes, a dense sampling would be needed. However, this is contrary to the goal of reduced sampling. Further, spatial upscaling typically averages parameters within spatial blocks and short scale variability is lost. With the approach proposed in this study, short scale processes are still represented well while aggregating spatially. Unfortunately, this process didn't work well for K_s which is a critical need. Some have found a link between K_s and texture (Jaynes and Tyler, 1984; Puckett et al., 1985) while others have found poor predictions from texture even when including bulk density and SOC (Weynants et al., 2009). Pachepsky et al. (2006) and Li et al. (2006) suggest using soil structural information as predictors for K_s. Perhaps inclusion of structural variables in the method proposed here could improve prediction zones for distinguishing K_s groups. However, addressing the large uncertainties associated with inversion modeling of K_s could prove beneficial as well.

The concept of using spatial statistics as a tool to make predictions about soil hydraulic properties has been demonstrated in this work. Using FKA, underlying causal processes can be identified

from soil physical properties and possibly just from exhaustive data. Once underlying processes are mapped, they can be used to define zones of differing properties. This is a first step toward developing a spatially based PTF. However, after determining the zones, characterizing the hydraulic curve that represents each zone is still required. Future work should investigate the use of regionalized factors to "back calculate" to variables of interest or develop a regionalized structural equation modeling scheme such as partial regression. This would be the spatial version of using a regression line to predict dependent variables from independent variables. Also, investigating the spatial relationships related to K_s in more detail may yield new information about this parameter. FKA has been used by many to identify underlying processes and define management zones. Yet, there is still much power left in this technique waiting for discovery and development.

Citations

Alary, C. and H. Demougeot-Renard. 2010. Factorial kriging analysis as a tool for explaining the complex spatial distribution of metals in sediments. *Environmental Science Technology*. 44: 593-599.

Bezdek, J.. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY: Plenum Publishing Company.

Biswas, B. and B. Si. 2009. Spatial relationship between soil hydraulic and soil physical properties in a farm field. *Canadian Journal of Soil Science*. 89: 473-488.

Bocchi, S., A. Castrignanò, F. Fornaro, and T. Maggiore. 2000. Application of factorial kriging for mapping soil variation at field scale. *European Journal of Agronomy*. 13: 295-308.

Botros, F., T. Harter, Y. Onsoy, A. Tuli, and J. Hopmans. 2009. Spatial Variability of Hydraulic Properties and Sediment Characteristics in a Deep Alluvial Unsaturated Zone. *Vadose Zone Journal*. 8(2):276-289.

Ciollaro, G. and N. Romano. 1995. Spatial variability of the hydraulic properties of a volcanic soil. *Geoderma*. 65: 263-282.

Chilés, J. P. and P. Delfiner. 2012. *Geostatistics: Modeling Spatial Uncertainty 2nd ed*. Hoboken, NJ: John Wiley & Sons, Inc.

Corwin, D. and S. Lesch. 2005. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*. 46: 11-43.

Dane, T. and G. Topp. eds. 1994. *SSSA Book Series: 5, Methods of Soil Analysis, Part 4 Physical Methods*. Madison, WI: Soil Science Society of America.

Deurer, M. and W. Duijnisveld. 2000. Spatial analysis of water characteristic functions in a sandy podzol under pine forest. *Water Resource Research*. 36(10): 2925-2935.

Dobermann, A., P. Goovaerts, and T. George. 1995. Sources of soil variation in an acid Ultisol of Philippines. *Geoderma*. 68: 173-191.

Doussan, C. and S. Ruy. 2009. Prediction of unsaturated soil hydraulic conductivity with electrical conductivity. *Water Resource Research*. 45: W10408, doi:10.1029/2008WR007309.

Fridgen, J., N. Kitchen, K. Sudduth, S. Drummond, W. Wiebold, and C. Fraisse. 2004. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agronomy Journal*. 96: 100-108.

Gambolati, G., M. Putti, P. Teatini, and G. Gasparetto Stori. 2006. Subsidence due to peat oxidation and impact on drainage infrastructures in farmland catchment south of Venice Lagoon. *Environmental Geology*. 49: 814-820.

Glass, V., P. Peckham, and J. Sanders. Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*. 42(3): 2377-288.

Goovaerts, P.. 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*. 43: 597-619.

Grossman, R. and T. Reinsch. 2002. Bulk density and linear extensibility. In Dane, J.H., and G.C. Topp. Eds. *Methods of soil analysis, Part. 4*. Madison, WI: Soil Science Society of America.

Hammel, K., J. Gross, G. Wessolek, and K. Roth. 1999. Two-dimensional simulation of bromide transport in a heterogeneous field soil with transient unsaturated flow. *European Journal of Soil Science*. 50: 633-647.

Jaynes, D. and E. Tyler. 1984. Using soil physical properties to estimate hydraulic conductivity. *Soil Science*. 138: 298-305.

Li, Y., D. Chen, R.E. White, A. Zhu, and J. Zhang. 2006. Estimating soil hydraulic properties of Fengqiu Country soils in the North China Plain using pedo-transfer functions. *Geoderma*. 138: 261-271.

Lin, Y., Y. Tan, Y. Lin, C. Liu, and C. Hung. 2004. Geostatistical method to delineate anomalies of multi-scale spatial variation in hydrogeological changes due to ChiChi earthquake in the ChouShui River alluvial fan in Taiwan. *Environmental Geology*. 47: 102-118.

Mallants, D., M. Binayak, J. Diederik, and J. Feyen. 1996. Spatial variability of hydraulic properties in a multi-layered soil profile. *Soil Science*. 161(3): 167-181.

Mallants, D., P. Tseng, N. Toride, A. Timmerman, and J. Feyen. 1997. Evaluation of multimodal hydraulic functions in characterizing a heterogeneous field soil. *Journal of Hydrology*. 195: 172-199. (A)

Mallants, D., B. Mohanty, A. Vervoot, and J. Feyen. 1997. Spatial analysis of saturated hydraulic conductivity in a soil with macropores. *Soil Technology*. 10: 115-131. (B)

Mertens, J., D. Jacques, J. Vanderborght, and J. Feyen. 2002. Characterisation of the field-saturated hydraulic conductivity on a hillslope: in situ single ring pressure infiltrometer measurements. *Journal of Hydrology*. 263: 217-229.

Mualem, Y. 1976. A New Model for Predicting the Hydraulic Conductivity of Unsaturated Conductivity of Unsaturated Porous Media. *Water Resources Research*. 12(3): 513-522

Nanos, N. and J. Martín. 2012. Multiscale analysis of Heavy metal contents in soils: Spatial variability in Duero river basin (Spain). *Geoderma*. 189-190: 554-562.

Odeh, I., A. McBratney, and D. Chittleborough. 1992. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. *Soil Science Society of America Journal*. 56: 505-516.

Pachepsky, Y.A., W.J. Rawls, and H.S. Lin. 2006. Hydropedology and pedotransfer functions. *Geoderma* 131:308–316.

Puckett, W., J. Dane, B. Hajek. 1985. Physical and mineralogical data to determine soil hydraulic properties. *Soil Science Society of America Journal*. 49: 831-836.

Radcliffe, D. and J. Simunek. 2010. *Soil Physics with HYDRUS: Modeling and Applications*. Boca Raton, FL: Taylor and Francis Group.

Roubens, M. Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*. 10: 294-301.

Russo, D. and W. Jury. 1987. A Theoretical Study of the Estimation of Correlation Scale in Spatially Variable Fields 1. Stationary Fields. *Water Resource Research*. 23(7): 1257-1268.

Scudiero, E., P. Teatini, D. Corwin, A. Berti, R. Deiana, F. Morari. Submitted. Delineation of site-specific management units in a saline area at the southern margin of the Venice Lagoon, Italy. *Computers and Electronics in Agriculture*.

Simunek, J., M. Sejna, H. Saito, M. Sakai, and M. Th. van Genuchten, The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.0x, *Hydrus Series 3,* Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA, 2008.

Sollitto, D., M. Romic, A. Castrignanò, D. Romic, H. Bakic. 2010. Assessing heavy metal contamination in soils of Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena*. 80: 182-194.

USDA Soil Survey Staff. 1998. *Keys to Soil Taxonomy 8th ed*. Washington, D.C.: USDA.

van Genuchten, M. 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soil. *Soil Science Society of America Journal*. 44(5): 892-898.

Vereecken, H., R. Kasteel, J. Vanderborght, and T. Harter. 2007. Upscaling Hydraulic Properties and Soil Water Flow Processes in Heterogeneous Soils: A Review. *Vadose Zone Journal*. 6: 1-28.

Vereecken, H., M. Weynants, M. Javaux, Y. Pachepsky, M.G. Schaap, and M.Th. van Genuchten. 2010. Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: A Review. *Vadose Zone Journal*. 9: 795-820.

Wackernagel, H.. 2003. *Multivariate Geostatistics 3rd ed*. Berlin, Heidelberg, Germany: Springer-Verlag.

Weynants, M., H. Vereecken, and M. Javaux. 2009. Revisiting Vereecken Pedo transfer functions: Introducing a Closed-Form Hydraulic Model. *Vadose Zone Journal*. 8:86-95.

Wierenga, P., R. Hills, and D. Hudson. 1991. The Las Cruces Trench Site: Characterization, Experimental Results, and One-Dimensional Flow Predictions. *Water Resource Research*. 27(10): 2695-2705.

Chapter 4: Conclusion

Summary

Current efforts in soil science research are aimed at a better understanding of unsaturated hydraulic properties of soil. Many fields including precision agriculture and climate sciences could benefit from an improved understanding of water movement in the vadose zone because this zone is a critical interface between the atmosphere, terrestrial vegetation and groundwater. However, this understanding is limited by unreliable measurements, scale effects and the need for intrusive, expensive data collection techniques. While many researchers have investigated these problems, there is limited literature involving the application of geospatial statistics. Geospatial statistics includes information about location in analysis and allows for another avenue in which relationships may be explored. In this thesis research, multivariate geospatial statistics were applied to describe the spatial distributions, explore relationships and make predictions of soil hydraulic parameters. In particular factorial kriging analysis (FKA) was used to distinguish scale dependent relationships and identify the underlying processes responsible for the spatial variation of hydraulic properties.

In one study, FKA was applied to both hydraulic parameters and soil physical properties to explore the spatial relationships. The hydraulic parameters are the parameters from the van Genuchten (1980) and Mualem (1976) models for unsaturated water retention curves (WRC) and unsaturated hydraulic conductivity curves (UHC) respectively. Relationships between hydraulic parameters and other hydraulic parameters as well as hydraulic parameters with soil physical properties were examined. Some relationships are already well known but some are also new. One example is a complex interaction

between K_s, n, bulk density, and texture where the interaction of bulk density and sand is related to the direction of the relationship between K_s, and n. Applying FKA allows for mapping of the regionalized PCs which represent sources of variation. These sources of variation may be interpretable as the underlying processes responsible for the variation. For this study, short scale ranges were around 105m while long scale ranges were around 235m. Bulk density and the first short scale PC were almost identical. The distribution of the first PC at long scale resembled the paleo-channel. Another interesting finding was the relative magnitude between short scale and long scale variance as a function of depth. Near the surface, short scale PCs were much larger than long scale. However, as depth increases, long scale PCs become more dominant. This is likely because of the increased activity near the surface.

In a second study, FKA was applied to hydraulic parameters and soil physical properties separately. The retained PCs were used to define zones by applying the fuzzy-c clustering algorithm. The zones defined from the soil physical properties were compared to those defined by the hydraulic models to explore the potential for use as a predictive tool. Between 93.35% and 46.22% of the nodes were placed in the same groups yielding between a 152.34% and 39.81% increase over a random selection. Further examination of the predictive capability involved plotting and comparing the known hydraulic curves as sorted by the zones. This analysis revealed that the zones were best at distinguishing between WRC curves mostly because of accurate separation of θ_s. Se curves were distinguished fairly well while UHC curves were not sorted well. The failure to distinguish UHC curves is mostly due to a failure to distinguish K_s.

Future Work

While this work has made progress, there are still many opportunities to improve. With regards to exploring spatial relationships, many more sites and soil types need to be examined before clarity can be reached. The site used in this study has a wide range of soil properties but the relationships found

may be very different at other sites. Without more data to confirm or bound the relationships present in this site no vast, soil-wide conclusions can be drawn. With more study sites and soil types, better models and understanding will become attainable. This study also developed a method that could allow for prediction of hydraulic zones which could be used to direct sampling and reduce the number of hydraulic samples needed to capture the majority of the variance in the field. This reduction in needed samples may allow for larger study sites to be investigated. However, the method did not perform well for UHC. One possible solution would be to remove θ_s and θ_r from the spatial models so that the PCs would focus on just the parameters that affect UHC. Subsequent investigation of the descrepencies between the two models may reveal new information about Ks and its relationship with other soil properties. Alternatively, an approach similar to chapter 2 could be applied to a variable set which excludes θ_s and θ_r thereby focusing the study on UHC.

In particular to this site, more short range samples should be taken. Many of the hydraulic parameters in this study are known to have ranges shorter than the smallest sample separations used in this study. While the overall goal is to study the longer range variability, verifying that there is a very short scale structure will reduce the nugget effect. By reducing the nugget effect, more confidence in data quality (repeatability and scale effects) will be gained.