8-2015

Geostatistical Analysis of Point Soil Water Retention Parameters for Flint Sand

William Blake Roberts
University of Tennessee - Knoxville, wrober26@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by William Blake Roberts entitled "Geostatistical Analysis of Point Soil Water Retention Parameters for Flint Sand." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geology.

Edmund Perfect, Major Professor

We have read this thesis and recommend its acceptance:

Larry McKay, John Tyner

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Geostatistical Analysis of Point Soil Water Retention Parameters for Flint Sand

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

William Blake Roberts
August 2015

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for his blessings throughout my academic career and personal life. I would also like to thank my family and friends for their unending encouragement and support through my education. I would like to dedicate my thesis to my late grandpa, Harlan Roberts. He was a man's man that touched the lives of everyone around him through his kind heart and unrivaled work ethic. He was proud to call me his grandson, and because of that, I will continue to work hard every day to honor his legacy.

I would like to express my sincere gratitude to my thesis advisor, Dr. Ed Perfect. It was a great privilege and honor to work and study under his guidance. I would like to thank him for giving me the opportunity to obtain my Master's degree from the University of Tennessee, Knoxville (UT). I would also like to thank Brendan Donnelly for his help and friendship throughout our time working as graduate students under Dr. Perfect. Thanks also to my committee members, Dr Larry McKay and Dr. John Tyner, for their help throughout this process. I would also like to thank the entire Earth and Planetary Sciences Department for making my experience at UT the most enjoyable time of my life!

ABSTRACT

Geostatistics were employed to characterize sub-core scale heterogeneity and identify spatial structure in previously published water retention data (Kang et al., 2014) obtained using neutron radiography for Flint sand. The water retention data were parameterized using the Brooks and Corey (BC) model. The BC parameters investigated were: saturated water content (θ_s), residual water content (θ_r), air entry value (ψ_a), and pore size distribution index (λ). Spatial dependency in the BC parameters was identified using semivariograms. Of the four BC parameters analyzed, two were found to be spatially correlated, θ_s and ψ_a. The spherical model fit to the cross variogram was used to perform co-kriging and map out the spatial dependency of these parameters. Low and high values apparent at the top and bottom of the kriged map for ψ_a implicated packing and compressive stress as the major causes of sub-core scale heterogeneity for this parameter. A concentrated area of high values in the center of the kriged map for θ_s suggests that neutron scattering and the normalization procedure employed during image analysis to eliminate the effect of variable neutron path lengths was not completely successful. To alleviate these effects a trend correction process was developed by generating a second dataset using cross-validation, calculating the difference between the observed and leave-one-out cross validation data set, and adding the average of the observed data to the newly created residual variable. This trend correction process was validated using an independent data set collected by Cropper (2014). Mann-Whitney and Kolmogorov-Smirnov two sample tests were employed to determine if the Cropper (2014) parameters were significantly different from the trend corrected parameters in terms of their median values and frequency distributions, respectively. The results from both tests found

significant differences between the two data sets indicating the trend correction procedure was unsuccessful, likely due to the unconsolidated sample and cylindrical geometry employed. Since spatial structure can have profound effects on flow and transport predictions, future work using neutron radiography to measure point BC parameters should focus on consolidated samples and rectangular sample geometry. Further exploration of the novel trend correction procedure is warranted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Problem Statement

Spatially correlated heterogeneity occurs at multiple scales and represents a major challenge for predicting fluid movement in the subsurface (Pini and Benson, 2013). Investigation of multi-scale heterogeneity and its effects on flow and transport is important for multiple applications such as site characterization, contaminant transport, brine displacement efficiency, CO_2 sequestration, and reservoir production (Duijn et al., 1995**;** Eaton, 2006; Perrin and Benson, 2009; Li and Benson, 2014). Simulation studies have shown that sub-core scale heterogeneity can have a profound effect on flow and transport behavior (Pini and Benson, 2013). Li et al. (2013) presented evidence indicating that even small inaccuracy in characterizing multiphase flow parameters can accumulate and lead to significant error in long-term modeling at larger scales. Therefore, continuing investigation of sub-core scale heterogeneity is essential to better understand and model fluid movement in the subsurface.

To understand and model multiphase flow in the subsurface the water retention curve is often employed. Kang et al. (2014) developed a technique for directly measuring water retention curves at the sub-core scale using 2-D neutron radiography. This technique allows multiple curves to be calculated for a single core. Kang et al. (2014) computed 120 water retention curves for a single repacked core sample of homogeneous sand material. The water retention curves were parameterized by fitting the Brooks and Corey (BC) equation (Brooks and Corey, 1964) to each of the 120 curves. Histograms of the BC parameters show a range in distributions despite water retention curves being calculated

for a nominally homogenous material. Distributions within each parameter can be explained by sub-core scale heterogeneity caused in three ways:

1. natural heterogeneity which was effectively eliminated when the Kang et al. (2014) experiment was conducted on a nominally homogeneous material
2. artificial heterogeneity due to the measurement method itself (i.e. neutron imaging)
3. induced heterogeneity due to stratification caused by packing and particle segregation

Parameter histograms illustrate a distribution, but they do not characterize the spatial distribution of the parameters. Spatial heterogeneity is known to effect fluid flow and transport through porous media (Zinn and Harvey, 2003). Consequently, further investigation is needed to understand the spatial distribution of the BC parameters and identify possible causes of the sub-core scale heterogeneity in order to potentially improve modeling of flow and transport at this scale.

The purpose of this study is to apply geostatistical techniques to the BC parameters computed by Kang et al. (2014) to identify spatial structure and characterize sub-core scale heterogeneity of the repacked, nominally homogeneous sand material. Geostatistics is typically applied at the field scale. Very few studies have applied geostatistics at the sub-core scale, and it is unknown whether any previous studies have used geostatistics to identify heterogeneity of BC parameters at the sub-core scale. Understanding the role of heterogeneities at this scale provides a first step toward appropriate methods for up scaling laboratory-measured data to simulate field-scale multi-phase flow behavior (Perrin and Benson, 2009).

1.2 Objectives

The overall goals of this research are to characterize, explain, and attempt to correct method based sub-core scale heterogeneity of a repacked, nominally homogeneous sand material. In order to accomplish this, geostatistics will first be applied to BC parameters computed by Kang et al. (2014). Subsequently comparison statistics will be used to validate the results of the geostatistical analysis. The objectives of the research are:

1. Produce semivariograms of BC parameters measured by Kang et al. (2014) at Oak Ridge National Lab (ORNL)
2. Produce kriged maps of the BC parameters to evaluate how they vary across the sand column
3. Remove spatial structure caused by the data collection procedure in the Kang et al. (2014) dataset
4. Validate spatially-corrected neutron parameters against parameters collected using an independent methods

1.3 Hypotheses

The null hypothesis is that there will be little or no spatial structure in the measured Brooks and Corey parameters because they were fitted to water retention curves calculated for a nominally homogeneous sand material. The alternative hypothesis is that if there is spatial structure, it will be limited to the vicinity of the ends of the core and will be due to inconsistency in core packing.

CHAPTER 2
LITERATURE REVIEW

2.1 Sub-Core Scale Heterogeneity

Understanding sub-core scale heterogeneity is important because it can influence fluid movement at much larger scales. The presence of sub-core scale features can significantly increase overall variability and connectedness in the subsurface (Zinn and Harvey, 2003). In this section, several studies will be investigated to illustrate specific examples of how sub-core scale heterogeneity influences fluid movement. Also, these studies demonstrate the importance of research dedicated to better understanding and improving our ability to model sub-core scale heterogeneity.

In the study conducted by Pini (2012), CO_2/water drainage capillary pressure saturation curves were measured at reservoir pressure and temperature for the Berea sandstone. Its core was initially saturated with water and then flooded with 100% CO_2 at increasingly high flow rates, X-ray Computed Tomography (CT) images were captured each time the flow rate was increased (Pini, 2012). This technique allowed observation of capillary-pressure saturation relationships on mm-scale subsets of the rock core (Pini, 2012). Results from this study show that sub-core scale capillary-pressure saturation curves are different from one another, this suggests that there is a significant degree of heterogeneity in the supposed homogeneous Berea sandstone core (Pini, 2012). Also, the study shows that sub-core scale heterogeneities can cause small changes in capillary pressure which lead to strong variations in saturation.

Krause et al. (2013) conducted steady-state relative permeability measurements at reservoir conditions. X-ray CT scanning was used to measure sub-core scale porosity and saturation distributions (Krause et al., 2013). The work was conducted on two rock cores: Berea sandstone, which in this study was considered to have a low degree of heterogeneity, and a sandstone core from the Waare C formation which had a high degree of heterogeneity. Results from the study show that the Berea sandstone had a homogeneous porosity distribution, but due to sub-core scale heterogeneities the permeability distribution was strongly heterogeneous. Considerable heterogeneity was found in both the porosity and permeability distributions for the Waare core. This method was shown to accurately characterize sub-core scale heterogeneity, and provide well-characterized rock core for applications such as upscaling (Krause et al., 2013).

Pini and Benson (2013) characterized the core/sub-core scale heterogeneity of a naturally heterogeneous sandstone using N_2/water core flooding, X-ray CT imaging, and a scaling factor. They developed a procedure that can potentially combine pore and reservoir scale flow properties while sufficiently characterizing sub-core scale heterogeneities. Pini and Benson (2013) also acknowledge that geologic heterogeneity occurs over a wide range of scales and that it presents a major challenge when attempting to predict fluid movement in the sub-surface.

Li and Benson (2014)stated that, “spatially correlated heterogeneities at small scales can affect large-scale CO_2 migrations in geological sequestration.” The purpose of their work was to show the effect of small-scale “High-Randomness Pattern-Free” (HRPF) heterogeneities on buoyancy driven CO_2 migration. The study was conducted on a 2D

permeability field (5m x 0.4m) that was relatively homogenous over the entire sample domain. However, it was shown that heterogeneities were found at the millimeter scale. These heterogeneities were determined to be highly random because a short spatial correlation was shown, using a variogram, at the millimeter scale. Because millimeter scale heterogeneity appears to be relatively homogenous at a larger scale, it is often ignored in core analysis and flow simulation (Li and Benson, 2014). It was shown that CO_2 plume velocity can be significantly decreased or even completely immobilized by millimeter scale heterogeneity if the plume volume is small.

2.2 Computed Tomography and Neutron Imaging

Neutron imaging was first applied in soil science in the 1970's to visualize the spatial distribution of soil water characteristics (Perfect et al., 2014). Neutron imaging is ideal for observing air-water displacements in soil because neutrons are strongly attenuated by hydrogen rich fluids but not by air or soil minerals. Deinert et al. (2004) concluded that neutron imaging is ideal for detailed laboratory studies because of its sensitivity to variations in moisture content and its ability to image nontranslucent material (Cheng et al., 2012).

Tumlinson et al. (2008) first applied neutron imaging to measure a point water retention curve. Their analysis was based on a single applied matric potential; therefore, it was not possible to construct a full drainage function. Papafotiou et al. (2008) determined the 3-D distribution of water in a porous medium under quasi-steady state flow after two drainage steps. They also tested the ability of neutron and synchrotron tomography to determine average hydraulic properties using numerical simulations.

Vasin et al. (2008) conducted multi-step drainage experiments in two heterogeneously packed sand columns to test the influence of packing structures on the movement of water and to assess outflow curves in both columns made with an upscaled model. Neutron radiography (2-D) was used to measure the spatial distribution of water content during the multi-step drainage process and neutron tomography was used to measure 3-D water content at the steady state. Cheng et al. (2012) demonstrated that neutron imaging can fully reproduce the hysteretic characteristics of soil water and can be used to quantify average relative saturations. Kang et al. (2014) employed neutron radiography to map out the spatial distribution of water and determine pixel (1 pixel = 0.0032 cm or 32 μm) or point water retention functions at different locations within a nominally homogenous Flint sand column. An 8 x 15 Cartesian grid was superimposed over each neutron radiograph. A water retention curve was calculated for each point on the grid system. Histograms of the water retention curve parameters indicate the existence of sub-core scale heterogeneities. Also, the presence of spatial structure in these data will influence upscaling strategies, and therefore a geostatistical analysis is warranted.

2.3 Geostatistical Analysis of Sub-Core Heterogeneities

Historically geostatistics was applied to estimate ore reserves in the mining industry (Matheron, 1963). However, variogram estimation and spatial prediction (kriging) have spread to all sciences where data exhibit spatial dependency (Ecker, 2003). Geostatistics has been applied to a wide range of disciplines and is defined by Olea (1991) as "the application of statistical methods … for use in the earth sciences, particularly in geology" (Ecker, 2003). Geostatistics is typically applied to data at the field scale (Dunlap and Spinazola, 1984; Iqbal et al., 2005; Huysmans and Dassargues, 2006; Saito et al., 2009);

however, Warrick et al. (1986) stated that "dimensionally, applications of geostatistics could be for distances of a few molecules or kilometers."

This section provides several examples of geostatistics being successfully applied at the sub-core scale. The success of geostatistics at this scale is important for the purpose of this study and its ability to effectively characterize sub-core scale spatial dependency. Commonly geostatistical methods applied at a small scale are associated with X-ray computed tomography (CT) scanning to identify core/sub-core scale heterogeneity. No examples were found of geostatistical analysis applied to sub-core heterogeneities quantified with neutron imaging.

Peters and Afzal (1992) were amongst the first to use geostatistics at the core-scale. They investigated core-scale heterogeneities of reservoir rock to access oil recovery processes. First, they conducted a CT-imaging experiment to characterize a laboratory sandpack. Then semivariograms, along with other statistical analyses, were used to access and visualize the variability and spatial continuity of porosity and permeability distributions (Peters and Afzal, 1992). It was determined that CT imaging is a very powerful tool for characterizing permeable materials. Results indicated that even though it is often assumed that a sandpack has homogenous and isotropic properties, sub-core heterogeneities resulted from packing. This phenomenon is commonly observed in the geotechnical literature (Oliviera et al., 1996; Bromly et al., 2007; Lewis and Sjöstrom, 2010) and is a feature that may be identified within this study. Porosity and permeability variations that resulted from this heterogeneity were successfully determined using geostatistics at the core scale.

Grevers and De Jong (1994) applied geostatistics at the sub-core scale to analyze the spatial continuity of compacted soil and subsoil macroporosity, obtained from CT images. Geostatistical analysis was useful to discriminate the spatial variability and continuity of soil macropores in subsoils at the micro-scale (i.e. less than 1 cm). Perret et al. (1998) developed a procedure to characterize macropore morphology in large undisturbed soil columns using a 4th generation X-ray CAT scanner. Geostatistics, along with other methods, were used to access the variations in macropore morphology as a function of depth. Semivariograms were computed at every 30 mm interval. Results from the geostatistical analysis indicate that the spatial variability of macroporosity decreased with depth.

Aylmore and Rasiah (1998) explored the use of geostatistics to clarify the issue of randomness and continuity on the spatial distribution of porosity, fractal dimension, water content, and unsaturated hydraulic conductivity at scales as small as 1-cm intervals, obtained using a custom-built scanner and computed tomography techniques. It was shown that even though randomness existed in the spatial distribution of the soil parameters, specific trends existed in their spatial continuity (Aylmore and Rasiah, 1998). The results indicate that geostatistical analysis is useful to clarify the issue of randomness at very small scales. Aylmore and Rasiah (1998) ended the study with a recommendation of using geostatistics to clarify randomness and spatial continuity of soil parameters, especially porosity because it can have a significant impact on flow and transport in porous media.

Murata and Saito (2003) used a bubble concrete material to create synthetic porous rocks with varying porosity from 10-60% to examine the influence of porosity and pore

size distribution on hydraulic properties. Binary images were taken of vertical and horizontal cross-sections of the synthetic rock samples. Thresholding of the images was performed to separate pores and matrix. Semivariograms were used to identify the spatial structure of the pore size distribution. Information about the spatial correlation provided by the semivariogram analysis was used to create three-dimensional simulations of pore distribution (Murata and Saito, 2003).

De Gryze et al. (2006) conducted a study to explore the effects of decomposing residue on porosity and pore morphology in artificial aggregates compared to native field aggregates. X-ray computed tomography (CT) was used to visualize the pore space in 3D. Mass fractal dimension and variogram analysis were performed on the images to quantify the micro-meter scale changes in porosity and pore morphology. Both types of soil samples were found to have significant spatial correlation, apparent by the standardized variograms increase from a small nugget to a sill and the successful fit of an exponential model (De Gryze et al., 2006). Although variogram analysis successfully identified spatial correlation for the soils a meaningful or interpretable difference in porosity and pore morphology between the artificial and native field aggregates caused by decomposing residue was not identified.

Houston et al. (2013) is another example of geostatistical techniques conducted on CT images at the sub-core scale. In this study indicator kriging is used to deal with image segmentation issues relevant to 3D CT images of naturally occurring heterogeneous materials such as soils. The method developed by Houston et al. (2013) is an extension of the commonly used indicator-kriging algorithm of Oh and Lindquist (1999). The results of

this study show that the proposed extension produce image segmentation results similar to that of the original method with a substantial decrease in cost (Houston et al. (2013). This paper indicates that adaptive window kriging is a viable solution to solving segmentation issues at micro-metric resolutions (Houston et. al, 2013).

Finally, Carducci et al. (2014) conducted a study aimed at evaluating the spatial and morphological configuration of the pore space in core samples of kaolinitic Latosols. 3D images of the core samples were generated using an X-ray CT scanner. To prepare the 3D images for geostatistical analysis they followed an image processing technique in which grey-scale images were converted into binary images, distinguishing the void and non-void space (Carducci et al., 2014). Semi-variograms were then computed to identify the spatial structure of the different soils. It was shown that greater spatial variability was apparent in the horizontal direction of the 3D images. The identification of spatial variability in agricultural soils is useful to monitor the effects agricultural systems have on pore distribution (Carducci et al., 2014).

These studies offer encouraging evidence that geostatistics can be used to identify spatial structure in the Kang et al. (2014) dataset. They also provide insight into a potential cause of sub-core scale heterogeneity (i.e. packing) within a nominally homogenous sandpack; as well as, recommend the use of geostatistics to identify the spatial structure of porosity. The previously mentioned studies typically relied on the binarization of CT images before using geostatistics. This research is unique in that geostatistics is applied to data evenly spaced on a Cartesian grid system at the sub-core scale, a method much more similar to how geostatistics is applied at the field scale.

CHAPTER 3
METHODOLOGY

Geostatistics is playing an increasingly significant role in hydrogeology and reservoir modeling. This increased role is driven by the need to visualize and quantify subsurface heterogeneity because of its control on groundwater flow, solute transport, and multiphase fluid movement. In this section, the process of applying geostatistics to identify sub-core scale heterogeneity in the Brooks and Corey water retention parameters estimated by Kang et al. (2014) is described. Subsequently, the correction process used to remove spatial dependency caused by the Kang et al. (2014) data collection procedure is described. The R statistical software (R Core Team, 2014) and SAS 9.2 (SAS Institute, Inc., 1996) programs were used to conduct the geo statistical analysis in this section.

3.1 Data

Neutron imaging combined with the hanging water column method were employed by Kang et al. (2014) to measure point (pixel-scale) water retention curves for repacked Flint sand. The hanging water column setup consisted of a cylindrical aluminum container (inner diameter = 2.56 cm, height = 10 cm) connected with Tygon tubing via an outlet at its base to a burette filled with distilled water. Oven-dried Flint sand (~ 50 g) was saturated with water and then incrementally packed into the Al container up to 5.6 (± 0.1) cm (Kang et al., 2014). Before the drainage experiment, the sand column was fully saturated with water by raising the water level in the burette to a height approximately equal to the top of the sand pack and allowed to equilibrate overnight (Kang et al., 2014).

Neutron imaging of the sand column was performed using cold neutrons at the High Flux Isotope Reactor (HFIR) CG1-D beam line at Oak Ridge National Lab (ORNL), photographs of the experimental setup are shown in Figure 1 (all figures and tables are located in the appendix). The sand column was subsequently drained stepwise under quasi-equilibrium conditions by adjusting the height of the hanging water column to give various basal matric potential values of -2.1, -8.0, -11.8, -15.0, -16.6, -18.3, -20.8, -25.3, and -46.7 cm (Kang et al., 2014). Radiographic images (Figure 2) were acquired at each quasi-equilibrium state during the drainage process with an exposure time of 60s.

A grid of 8 x 15 was superimposed on the images (Figure 2), and the individual volumetric water content values (θ) at these locations were combined with information on the known distribution of matric potentials (ψ) within the sample to give point water retention curves for the 120 selected locations (Kang et al., 2014). The point water retention data were parameterized by fitting the Brooks and Corey, BC, equation (Kang et al., 2014). The Brooks and Corey equation is given by:

$$\theta = \theta_s \qquad 0 < \psi \leq \psi_a$$
$$\theta = \theta_r + (\theta_s - \theta_r) \times \left[\frac{\psi_a}{\psi}\right]^{\lambda} \qquad \psi > \psi_a \tag{1}$$

where θ (cm^3cm^{-3}) is volumetric water content, θ_s (cm^3cm^{-3}) is the saturated volumetric water content, θ_r (cm^3cm^{-3}) is the residual volumetric water content, ψ_a (cm) is the air-entry value, and λ (dimensionless) is the pore-size distribution index. Equation (1) was fitted to the 120 point water retention data sets using segmented non-linear regression (Marquardt method) in SAS 9.2. All of the fits converged successfully, except for 5 points which were excluded from further analyses. The median (and mean) RMSE and R^2 values for the successfully converged fits were 0.019 (0.024) and 0.985 (0.975), respectively.

3.2 Preliminary Classical Statistics

Before applying geostatistics to the Kang et al. (2014) dataset each parameter was analyzed using classical descriptive statistical methods to obtain mean, median, variance, and standard deviation values. The correlation coefficient (r), coefficient of determination (R^2), and scatter diagrams were used to identify correlations between parameters. The descriptive statistics, along with histograms, were used to examine the frequency distribution of each parameter. The Shapiro-Wilk (SW) normality test (Shapiro and Wilk, 1965) was used to detect non-normal data and Quantile - Quantile (Q-Q) plots were employed to visualize the SW results. Non-normal parameters were transformed either using a logarithmic (i.e. $\log_{10}(x)$) or inverse (i.e. $1/x$) methods to obtain a normal or more normal distributions before applying geostatistical analysis. Transformed parameters were back-transformed before interpretation. Normality testing is essential because geostatistical methods are optimal when data are normally distributed, and significant deviations from normality can cause problems with spatial prediction (Bohling, 2005). These preliminary steps were performed in an effort to ensure quality results from the geostatistical analysis.

3.3 Trend Surface Analysis

To understand the main purpose of conducting a trend surface analysis some basic restrictions of geostatistics must first be briefly discussed. In order to perform geostatistics a weak form of stationarity known as the intrinsic hypothesis must be satisfied (Journel and Huijbregts, 1978). The intrinsic hypothesis as described by Vieira et al. (2010) requires that the mean and semivariance depend strictly on the separation distance of samples and

not on their coordinate location (Vieira et al., 2010). Often, experimental data will not satisfy this restriction because of an underlying trend; therefore, a trend removal procedure is warranted to perform geostatistics (Vieira et al., 1983). This necessitates the use of trend surface analysis to remove the trend. A trend surface is fit to the experimental data by the method of least squares; then the predicted trend surface is subtracted from the experimental data leaving only the residuals. Geostatistical analysis is then performed on the residuals, which satisfies the intrinsic hypothesis.

The first objective of the trend surface analysis is to visualize, in 2-dimensions, trends within the soil column for each of the BC parameters. This is accomplished by fitting a second degree polynomial equation to the experimental data using the linear model function in R (R Core Team, 2014). The second degree polynomial equation is given by:

$$\hat{z}_i = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy \quad (2)$$

where $\hat{z}_i$ is the estimated Brooks and Corey parameter, b_0 is the y-intercept, b_1 and b_2 are first degree coefficients of the slope of the x and y coordinates, b_3 and b_4 are second degree coefficients corresponding to the curvilinear trends of x^2 and y^2, and b_5 is the coefficient for the cross product term, $x \times y$. R uses Ordinary Least Squares (OLS) to estimate the slope coefficients (b_i), and fit the polynomial to the experimental data. Once the trend surface is fitted to the data the following statistics are available to determine goodness of fit and significance of the trend surface: coefficient of determination (R^2), F-statistic, probability associated with the F-statistic ($p > F$), residual degrees of freedom (Residual DF), model degrees of freedom (Model DF), standard error (SE), t-value, and probability associated with the t-value ($p > | t |$).

OLS fits a regression plane to the experimental data by minimizing the residuals (the vertical distances to the plane). In other words, OLS attempts to find a "best fit surface" that minimizes the distance between the experimental data points and the predicted surface. OLS fits the experimental data to the polynomial so that the sum of squared residuals (SSR) is minimized, i.e.:

$$SSR = \sum_{i=1}^{n} [\, z_i - \hat{z}_i]^2 \tag{3}$$

where z_i is the actual value and $\hat{z}_i$ is the predicted value. OLS is an appropriate approach when trying to fit a polynomial trend to spatial data; however, it is not a good interpolation technique because it does not assume residuals are spatially correlated (Dubrule, 2003). Trend surface analysis is a global interpolator that uses polynomial regression to fit a single function to all points in the study area. Interpolation using geostatistics employs information provided from a local, spatially correlated neighborhood. The variogram (discussed in detail later) is used to describe the spatial structure within the neighborhood and assign weights to measured data to make predictions. To make a distinction between these interpolation techniques, trend surface analysis identifies a global spatial trend while geostatistics identifies local spatial trends based on autocorrelation.

There are two questions to pose when evaluating how well the model describes the experimental data. First, how well does the trend surface fit the experimental data? This is typically referred to as goodness of fit and is measured using R^2 (Neiman, 2007). There are three types of variation which are of interest when describing the goodness of fit: the aforementioned SSR, the total sum of squares (SST), and the model sum of squares (SSM). SST is the total variation in the experimental data, and is computed by:

$$SST = \sum_{i=1}^{n} [\, z_i - \bar{z}_i]^2 \tag{4}$$

where, $\bar{z}_i$, is the mean of the experimental data. SSM is the amount of variation accounted for by the model, and is calculated by:

$$SSM = \sum_{i=1}^{n} [\, \hat{z}_i - \bar{z}_i]^2 \tag{5}$$

The R^2 statistic is used to summarize how well the experimental data predicted from the model match the actual values. It can be calculated as:

$$R^2 = 1 - \frac{SSR}{SST} \tag{6}$$

Second, is the trend surface statistically significant (i.e. does the model account for more variation than can be expected by chance)? The statistical significance of the trend surface can be measured using the F-statistic, and its associated p-value. The F-statistic is calculated by:

$$F - statistic = \frac{Model\ variance}{Error\ variance} \tag{7}$$

where Model variance is calculated by dividing SSM by the Model degrees of freedom (DF), and the Error variance is calculated by dividing SSR by the Residual DF. If the achieved F-statistic is greater than or equal to the F-statistic obtained purely by chance, the model significantly predicts the underlying trends in the experimental data (Neiman, 2007). The significance level for the probability associated with the F-statistic was set at 0.05 (i.e. 95% confidence level) and was used to determine whether the achieved F-statistic, under the null hypothesis that all model coefficients are zero, could have arisen from sampling error

alone (Neiman, 2007). If the p-value associated with the F-statistic is less than or equal to 0.05 the null hypothesis is rejected.

Once the F-statistic and associated probability confirm that the overall model has a significant trend the t-value is used to determine the significance of specific trend directions. The t-value is calculated by:

$$t - value = \frac{b_i}{SE} \tag{8}$$

where b_i is the slope coefficient and SE is the standard error. The significance level for the probability associated with the t-value was also set at 0.05 and was used to determine whether a particular slope coefficient is significantly different from zero. If the p-value associated with the t-value is less than or equal to 0.05 the specific trend direction/component is significant and contributes to the overall trend surface. The identification of specific trend directions is necessary for a specific form of geostatistics, known as Universal Kriging, which will be discussed later.

3.4 Detrending

Before the experimental data can be adequately analyzed using geostatistics the intrinsic hypothesis must be satisfied by trend removal (Vieira et al., 1983). Fortunately, there is a very simple and effective way of detrending the experimental data (Vieira et al., 2010). As previously mentioned, a trend surface is fitted to the data using OLS in order to identify the trend directions that significantly contribute to the overall trend surface. Next, a new variable, known as the residual variable, is constructed by subtracting the value of the trend surface function from the experimental data (Vieira et al., 2010):

$$z_{res} = z_i - \hat{z}_i \tag{9}$$

The newly calculated z_{res} is then analyzed again to ensure that the trend removal was successful.

By removing the trend there is a change in scale for the data. Because the trend surface describes the overall spatial variation for the variable, z_{res} will produce a new surface which will contain simply the remaining surface roughness and its values will fluctuate above and below zero (Vieira et al., 2010). If the degree of the trend surface sufficiently describes the underlying trend of the parameter the mean value of the residuals should be zero (Vieira et al., 1983). After detrending, the geostatistical analysis will be performed on the residual variable; the trend surface should be added back for the final steps in the geostatistical process.

3.5 Spatial Interpolation

3.5.1 Introduction

Geostatistical methods are well documented (Isaaks and Srivastava, 1989; Cressie, 1993; Goovaerts, 1997), but for the purpose of better understanding the spatial interpolation techniques to be used in this study, a brief account of the relevant methods (Kumar and Remadevi, 2006) will be given here. Geostatistics is based on the theory of the regionalized variable, defined by Materon (1971) as a property or process that spreads in space and exhibits a certain spatial structure. In order to make interpolations and adequately describe the processes controlling the regionalized variable, geostatistics is employed. Geostatistics uses the empirical semivariogram to create a model for spatially correlated variables to interpolate unsampled locations using a technique known as kriging (Hartkamp et al., 1999). The gstat package in R (Pebesma, 2004) was used to conduct the

geostatistical analysis in this study. The program can compute an empirical semivariogram, fit the semivariogram with a theoretical model, estimate values at unmeasured locations using several forms of the kriging equation (i.e. simple, ordinary, universal, co-kriging), and validate the kriging results using "leave-one-out" cross-validation.

3.5.2 Empirical Semivariogram and Cross - Variogram

The empirical semivariogram is a central concept in geostatistics. It models the spatial structure of the regionalized variable and provides weighting information to the kriging algorithm for interpolation. The semivariogram is of the form:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (10)$$

where $\gamma(h)$ is the estimated semivariance for the lag distance h; $N(h)$ is the number of experimental pairs separated by vector h; and $z(x_i)$ and $z(x_i + h)$ are the values of variable z at positions x_i and $x_i + h$, respectively. Huysmans and Dassargues (2006) provide a simple explanation of the semivariogram, "as the sample semivariance described as a function of spatial separation." Small semivariance values indicate a high degree of spatial correlation between parameter values separated by the lag vector; whereas, large semivariance values indicate a low degree of spatial correlation (Huysmans and Dassargues, 2006).

In order to describe spatial relationships between parameters, the cross-variogram is used. The cross-variogram enables the characterization of spatial dependency between two variables. For locations where measurements of two parameters *u* and *v* have been made, the cross variogram can be calculated using the following equation (Oliver, 2010):

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [\, z_u(x_i) - z_u(x_i + h)]\, [z_v(x_i) - z_v(x_i + h)] \tag{11}$$

where $\gamma_{uv}(h)$ is the cross variogram comparing parameters *u* and *v,* and *h* is a lag distance separating two locations x_i and $x_i + h$ (Papritz et al., 1993). The cross variogram is a measure of cross variability between two different parameters, and it addresses whether the change in one parameter is spatially similar to another (Lloyd, 2014). To compute the cross-variogram both parameters must be correlated and have common measurement locations (Oliver, 2010). In this study, all BC parameter values were calculated at the same locations using a superimposed grid on the neutron radiographic images.

3.5.3 Semivariogram Model

Fitting an appropriate theoretical model to the empirical semivariogram is another important step in geostatistical analysis. Three of the most commonly used models, spherical, exponential, and Gaussian, were fitted to the empirical semivariogram of each BC parameter. These theoretical models take the following mathematical form (Orjuela-Matta et al., 2012) :

$$\text{Spherical: } \gamma(h) = \begin{cases} c_o + c_1 * \left(\frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) & h \le a \\ c_o + c_1 & h > a \end{cases} \tag{12}$$

$$\text{Exponential: } \gamma(h) = c_o + c_1 * \left(1 - \exp\left(\frac{-3h}{a}\right)\right) \tag{13}$$

$$\text{Gaussian: } \gamma(h) = c_o + c_1 * \left(1 - \exp\left(\frac{-3h^2}{a^2}\right)\right) \tag{14}$$

where $\gamma(h)$ represents the theoretical semivariogram dependent on the lag distance *h*, *a* represents the range or decorrelation length, c_1 and c_o represent the partial sill and nugget, respectively. The gstat package in R (Pebesma, 2004) uses non-linear regression in

OLS to fit the theoretical model to the empirical semivariogram using an iterative process where the range, partial sill, and nugget are determined. The range measures the limit of spatial dependence for each BC parameter and is the distance at which the variogram reaches its sill (c_o+ c_1). The sill represents the point at which the semivariogram levels-off. This plateau occurs because at some lag distance points being compared are so far apart they are no longer related to one another; at this point their squared differences are equal to the variance around the average value (Davis, 2002). The nugget is the y-intercept of the variogram (Lakhankar et al., 2010). It can be caused by variability at very short distances for which no pairs of observations are available, sampling inaccuracy, or inaccuracy in the instruments used for measurement (Hartkanmp et al., 1999).

The goodness of fit of the theoretical model to the empirical data is determined using R^2 and Normalized Root Mean Square Error (NRMSE). The NRMSE is a non-dimensional form of the Root Mean Square Error (RMSE), and is computed by dividing the RMSE by the range of the observed data. The NRMSE is useful for comparing goodness of fit for parameters with different units. The theoretical model with the highest R^2 and lowest NRMSE was said to be the most appropriate model to represent the empirical semivariogram; this is referred to as the "best fit model." Later in the study, cross-validation will be performed to validate the best fit model selection by testing the predictive ability of each model. An optimal theoretical model will be chosen based on a collection of information gathered from this section and the cross-validation section.

Fitting of the semivariogram with a theoretical model is necessary for several reasons. First, the kriging algorithm requires input for all distances and directions, while the empirical semivariogram was calculated for specific distances and directions. Second, a

model provides a convenient parametric form of the semivariogram. Finally, an appropriate model is needed to ensure a positive definite covariance matrix, which is a requirement of kriging (Tang, 2005).

3.5.4 Kriging

After fitting a theoretical model to each empirical semivariogram, kriging is performed to interpolate BC parameter values at unmeasured locations using the gstat R package (Pebesma, 2004). Kriging is a generic term for several techniques (e.g. simple, ordinary, universal, and co-kriging) used to estimate spatially dependent data at unsampled locations based on a weighted linear sum of measured data (Majani, 2007). The weights used in the kriging algorithm are derived from the modeled semivariogram. Goovaerts (1997) states that, "all kriging estimators are but variants of the basic linear regression estimator *Z*(x)* defined as:"

$$Z^*(x) - m(x) = \sum_{i=1}^{n(x)} \lambda_i(x) \, [Z(x_i) - m(x_i)] \tag{15}$$

where x is a measured value inside a spatially related neighborhood around the estimation location x_i; $n(x)$ is the number of data points located inside the spatially related neighborhood used to make an estimation of $Z^*(x)$, i.e. only the closest points are significantly important; $m(x), m(x_i)$ are the expected mean values of $Z^*(x)$ and $Z(x_i)$, respectively; and $\lambda_i\ (x)$ is the kriging weight assigned to $z(x_i)$, a realization of the regionalized variable $Z(x_i)$ (Davis, 2002; Goovaerts, 1997).

All forms of kriging share the same objective, namely to determine the weights, λ_i, that minimize estimation variance under the unbiasedness constraint (Goovaerts, 1997). Shown mathematically as:

$$\sigma_E^2(x) = Var\{Z^*(x) - Z(x)\} \tag{16}$$

minimized under the unbiasedness constraint,

$$E\{Z^*(x) - Z(x)\} = 0 \tag{17}$$

The kriging form will differ based on the model adopted for the random function (RF) $Z(x)$, which can be broken down into a trend component, $T(x)$, and a residual component $R(x)$:

$$Z(x) = R(x) + T(x) \tag{18}$$

The residual component is modeled as a stationary RF with mean equal to zero and covariance function that depends only on lag, *h*, not position, *x* (Bohling, 2005):

$$\begin{gathered} E\{R(x)\} = 0 \\ Cov\{R(x), R(x+h)\} = E\{R(x) * R(x+h)\} = C_R(h) \end{gathered} \tag{19}$$

The residual covariance function, $C_R(h)$, is derived via information provided by the semivariogram model, i.e.

$$C_R(h) = Sill - \gamma(h) \tag{20}$$

therefore, the semivariogram model used in the kriging program will represent the residual component of the variable (Bohling, 2005). The expected value of the random variable *Z* at location *x* is thus the value of the trend component at that location (Goovaerts, 1997):

$$E\{Z(x)\} = m(x) \quad (21)$$

The three main kriging variants, simple, ordinary, and universal differ according to their treatment of the trend component, $m(x)$.

Simple kriging (SK) is mathematically the least complicated, as the name implies, and considers the mean, $m(x)$, to be known and constant throughout the study area (Goovaerts, 1997). This assumption is satisfied when dealing with residuals from a function fitted by least squares (i.e. trend surface residuals), in this case the mean is zero (Davis, 2002). A form of simple kriging, known as residual kriging (RK), will be utilized in this study to deal with non-stationary regionalized variables. In RK the trend is estimated (using trend surface analysis) then removed from the data. The empirical semivariogram and fitted model are computed for the residuals in the previously stated manner, and then the residuals are kriged to obtain estimates. Finally, the trend is then added back to the kriged estimates to produce the final result.

Ordinary Kriging (OK) differs from SK in that the requirement of a known mean is dropped. OK accounts for fluctuations in the mean by limiting the domain of stationarity to the local neighborhood (Goovaerts, 1997). Where SK assumes a stationary known mean, OK assumes a stationary but unknown mean; Universal Kriging (UK) is a further generalization of the kriging procedure that removes the restriction of a constant mean. UK performs in a simultaneous process what would otherwise consist of three arduous steps: trend removal, kriging of residuals, and adding back the removed trend to residual estimates (Davis, 2002). In the gstat package (Pebesma, 2004) the underlying geographic

trend of the parameter is first determined using trend surface analysis and then supplied to the UK algorithm.

Co-kriging is an extension of the basic kriging algorithm that allows information provided by a spatially correlated co-variable to supplement the estimation process of the primary variable. For example, if $n_j(x)$ saturated water content data, $Z_2(x_j)$, are available in addition to the $n_i(x)$ air entry data, $Z_1(x_i)$, the simple co-kriging estimate for the primary variable at any unsampled location *x* is:

$$Z^*(x) - m_1 = \sum_{i=1}^{n_i(x)} \lambda_i(x) \left[Z_1(x_i) - m_1\right] + \sum_{j=1}^{n_j(x)} \lambda_j(x) \left[Z_2(x_j) - m_2\right] \tag{22}$$

where m_1 and m_2 are the primary and secondary means assumed known and constant within the study area (Goovaerts, 1998). Kriging and co-kriging are theoretically identical and abide by the same unbiasedness and minimization requirements. The only difference being cokriging models four covariance functions as opposed to a single covariance function for kriging:

$$\begin{aligned} C_{11}(h) &= Cov\{Z_1(x),\, Z_1(x+h)\} \\ C_{22}(h) &= Cov\{Z_2(x),\, Z_2(x+h)\} \\ C_{12}(h) &= Cov\{Z_1(x),\, Z_2(x+h)\} \\ C_{21}(h) &= Cov\{Z_2(x),\, Z_1(x+h)\} \end{aligned} \tag{23}$$

where $C_{12}(h)$ is assumed to be identical to $C_{21}(h)$ (Goovaerts, 1997). It has been reported in several studies (Goovaerts, 1997; Hartkanmp et al., 1999; Majani, 2007) that co-kriging should be applied when the primary variable is under sampled. In this study all parameters have an equal amount of data points; therefore, co-kriging is used as an attempt to improve

prediction by incorporating information gained by the spatial correlation between different BC parameters.

3.6 Cross-validation

Cross-validation (CV) allows one to compare the effects of different semivariogram models and kriging techniques on interpolation results (Goovaerts, 1997). The method consists of selecting a theoretical model and associated parameters for the underlying semivariogram, removing one of the measured values from the data set, then re-estimating the removed value at that location by kriging with the selected theoretical model (Vieira et al., 2009). This is referred to as "leave-one-out" CV. This approach was used for several purposes in this study: first, validation of the theoretical semivariogram model selection; second, performance comparison for different kriging techniques (i.e. simple, ordinary, universal, co-kriging); and third, the suite of estimated values was used to compute detrended kriged maps (discussed further in the "Trend Correction" section).

The initial reason for employing leave-one-out CV was to confirm the model selection based on goodness of fit. CV achieves this by evaluating the predictive ability of the entire geostatistical model, which includes the semivariogram model and kriging form. Variations of the kriging method and semivariogram model are tested to determine which combination results in the most accurate predictions. There is a difference between how well a theoretical model fits the empirical data and how well the kriging estimator, using that model, interpolates values at unsampled locations. The goodness of fit is an indication as to which model will achieve the best predictions, but this relationship is not perfect; therefore, CV is used to validate the "best fit model" based on interpolation results.

In order to validate the selected semivariogram model (chosen based on the criteria listed in the previous section, i.e. R^2 and NRMSE) the kriging form is held constant while other theoretical models (with lower R^2 and higher NRMSE values) are tested. Once the best fit model selection is confirmed, the kriging variants are tested while keeping the theoretical model constant. An optimal geostatistical model is chosen based on all available information, goodness of fit and predictive ability. The predictive ability of a geostatistical model is determined by comparing interpolated and actual values; as well as, conducting a performance evaluation which directly compares kriging methods.

The difference between the estimated and known values (CV residuals) can be investigated using several statistical tools to assess predictive ability. The one-to-one graph of the CV residuals is commonly used and is computed by (Vieira, 2009):

$$Z^*(x_i) = a + b * Z(x_i) \quad (24)$$

where a is the intercept, b is the slope, $Z(x_i)$ and $Z^*(x_i)$ are measured and estimated values, respectively. Assuming the ideal scenario $Z(x_i)$ and $Z^*(x_i)$ would be identical, resulting in a correlation coefficient (r) and b of unity, and an a of zero, with the graph of $Z(x_i)$ and $Z^*(x_i)$ being a series of points on the one-to-one line (Vieira, 2009). If a (the intercept) is positive predicted values are always an overestimation of observed values, and if a is negative the inverse is true. If b (the slope) is less than unity then the estimator, $Z^*(x_i)$, is overestimating small values and underestimating large values, while if b is greater than the opposite is true. The quality of the estimation can be assessed by these parameters. Another very powerful statistical metric that can be derived from the one-to-one graph is NRMSE. In this

section NRMSE is employed to determine predictive ability, while it was previously used to evaluate goodness of fit.

A separate performance evaluation was conducted for the kriging methods by calculating percent prediction accuracy (% PA). The % PA was determined by directly comparing CV estimated values for the three kriging forms used in this study (i.e. OK/RK, UK and CK) to observed values. The number of times the absolute value of the difference between the CV estimated value and the actual value was smallest for a particular kriging method was summed, N_b, then divided by the total number, N_t, of measured values (i.e. 115) and multiplied by 100. Given in mathematical form by:

$$\%\ Prediction\ Accuracy\ = \frac{N_b}{N_t} * 100 \tag{25}$$

The sum of the computed %PA for each kriging form will be 100%. The optimal kriging estimator was determined using the following criteria: goodness of fit (i.e. R^2 and NRMSE), statistics derived from the one-to-one graph (a, b, r, and NRMSE), and %PA. By incorporating the information gained from several statistical criteria the most appropriate semivariogram model and kriging form was chosen to compute the final kriged map for each spatially dependent parameter.

3.7 Trend Correction

As previously mentioned in the CV methods section, the suite of CV estimated values was used to compute detrended kriged maps of spatially correlated BC parameters. This was done in an effort to remove any spatial correlation resulting from the data collection procedure conducted by Kang et al. (2014). To identify the true controls of flow and

transport at the sub-core scale, the Kang et al. (2014) dataset, hereafter referred to as the Kang data, must be devoid of procedural spatial structure. Failing to remove this type of spatial structure will result in misinterpretation of the true geologic controls on fluid movement. This section presents the methods used in order to develop a technique using geostatistics to remove non-geologic spatial structure.

The CV estimated values were used in the same manner as the Kang data to compute a kriged surface. The necessary steps are briefly presented here:

1. Preliminary statistics were used to summarize the CV data and test for normality
2. Non-normal data were transformed using an inverse transformation
3. Normally distributed data were used in a trend surface analysis to identify significant trends
4. Significant trends were removed by subtracting the trend surface function from the CV data (Vieira et al., 2010), this process created a new residual variable
5. An empirical semivariogram was computed for the detrended CV data
6. Theoretical models were fit to the empirical semivariogram and their fit was assessed using R^2 and RMSE
7. Cross-validation was used to validate the selected theoretical model
8. Finally, the same kriging method used to obtain the CV values was used to compute a kriged map of the CV data

Up to this point kriged maps have been computed for spatially correlated BC parameters and their CV estimated values. Now CV kriged estimates will be used to remove non-geologic spatial structure from the Kang data. The manner in which this was

accomplished is very similar to that of the detrending procedure already discussed. The difference between the Kang and CV kriged values is calculated. The computed residual value is then added to the average of the Kang data kriged predictions to create a new variable. This calculation is illustrated by:

$$Z_{ra} = (Z_k - Z_{cv}) + \bar{Z}_k \quad (26)$$

where Z_k and Z_{cv} represent Kang and CV kriged estimates, respectively; $\bar{Z}_k$ signifies the mean of Z_k; and Z_{ra} is the corrected BC parameter. The new variable, Z_{ra}, is then mapped to visually represent the now detrended BC parameter. In the next section this procedure will be validated using independent data.

3.8 Validation

In this section several nonparametric statistics and plots are presented in an attempt to validate the trend correction process, and explain obvious trends in the BC parameter kriged maps. The Mann-Whitney (MW) and Kolmogorov–Smirnov (KS) two sample tests were employed to validate the spatially-corrected BC parameters against BC parameters for Flint sand independently collected by Cropper (2014). In the Cropper (2014) study nine water retention curves were measured using the hanging water column method for Flint sand columns of various lengths (4.3 cm to 55 cm), and point BC parameter values were inversely computed from the measured water retention data using Trucell (Jalbert et al., 1999). Proving that the spatially-corrected data is not significantly different from the independent point data will indicate that the trend correction process was successful. The level of significance or alpha (α) value, used for hypothesis testing in this section was set at 0.05 (Harned, 1995) . If the test probability level was less than or

equal to $\alpha = 0.05$, then the null hypothesis that both groups have the same distribution is rejected (Harned, 1995). The statistical software pack used in this section was SAS 9.2 (SAS Institute, Inc., 1996).

The MW test is a nonparametric test that checks for equivalence of medians of two sample populations (Davis, 2002). This statistical test combines and ranks all values from the sample populations while paying no attention to which group individual values belong to. The smallest number is given a rank of 1, the largest number is ranked N, where N is the total number of values in the two sample populations. The MW test statistic is the sum of the ranks for observations from one of the samples, usually the smaller sample, and is denoted by *W*. Exact probabilities of occurrence for specified values of *W* have been worked out and are available in statistical tables for small sample sizes (Davis, 2002). For large sample sizes, *W* is approximately normally distributed and can be tested using the standard normal table to calculate probabilities (Davis, 2002).

The KS two sample test compares the cumulative relative frequency distributions of two data sets, and computes an α-value dependent on the largest absolute difference between the distributions. This maximum difference is the KS statistic, *D* (Davis, 2002). The KS test checks for any violation of the null hypothesis (i.e. differences in median, variance, or distribution); because of this, it has less power to detect shifts in the median but more power to detect changes in the shape of the distributions (Lehmann and D'Abrera, 2006). The α-value is used to determine the chance that the calculated KS *D*-statistic would be as large or larger than critical values of *D* (Davis, 2002). If the α-value is ≤ 0.05 it can be concluded that the two sample populations are drawn from different populations.

The Kruskal-Wallis (KW) test along with plots of the interquartile range (IQR) for each parameter, were used to validate the trends visually apparent in the kriged maps. The KW test is used to test the equivalency of several samples (Davis, 2002). In this case the samples are equivalent to the number of points along the x and y directions of the grid (8 X 15) superimposed over each radiographic image. There were eight points, k = 8, in the x direction, and 15 points, k = 15, in the y direction, both directions had unequal sample sizes. The KW test was performed in the x and y directions for each parameter resulting in a total of eight tests. The procedure of this test is very similar to that of the MW test, except it is extended to more than two samples. The observations from k samples are pooled and then ranked from smallest to largest. For each k group the sum of the ranks is found by:

$$R_k = \sum_{i=1}^{n_k} R(x_{ik}) \tag{27}$$

where $R(x_{ik})$ represents the rank of the ith observation in the kth sample (Davis, 2002). The total number of observations N is determined by:

$$N = \sum_{j=1}^{k} n_k \tag{28}$$

where n_k is the number of observations in the kth sample (Davis, 2002). From the sum of the ranks the KW H-statistic can be calculated:

$$H = \frac{12}{N(N-1)} \sum_{j=1}^{k} \frac{[R_k - n_k\,(N+1)/2]^2}{n_k} \tag{29}$$

H is approximately distributed as chi-squared (χ^2) with *k*-1 degrees of freedom; therefore, χ^2 can be used to evaluate probability. If significant differences were detected by the KW test, the IQR was plotted for graphical comparison to identify dissimilar distributions.

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Fitting the Brooks and Corey Equation

2-D neutron radiographic images of volumetric water contents at nine imposed basal matric potentials were obtained during drying of Flint sand at the HFIR CG1-D beamline. A total of 120 point soil water retention curves were obtained from these images as described earlier. The Brooks and Corey (BC), Equation (1), was fitted to each curve using segmented non-linear regression in SAS and 98.5% of the fits converged successfully yielding 115 sets of the four BC parameters. The RMSE of the fits ranged from 0.005 to 0.078 cm^3cm^{-3}, while the R^2 values (from predicted versus observed values) were between 0.860 and 0.999. The median (and mean) RMSE and R^2 values were 0.019 (0.024) and 0.985 (0.975), respectively.

4.2 Classical Descriptive Statistics

The best fit values of the 115 sets of point BC parameters (θ_s, θ_r, ψ_a, and λ) ranged between 0.30-0.55 cm^3/cm^3, 0.00-0.06 cm^3/cm^3, 14.50-19.10 cm, 3.55-17.59, respectively. These parameters were first summarized using classical descriptive statistics (Table 1) and histograms (Figure 3). The histograms reveal considerable sub-core scale variability in the repacked sand column possibly due to heterogeneity in local porosity, connectivity of pores, and pore-size distribution caused by the packing procedure employed (Kang et al. 2014). The θ_s parameter exhibited the greatest variation (Figure 3b, Table 1). The frequency distribution for θ_s is clearly positively skewed, with some very high estimates of θ_s. This is not surprising since θ_s is likely to be quite sensitive to packing, edge effects and

surface variations (Kang et al. 2014). Furthermore, it represents an estimate of the water content at saturation, which is the condition that is least accurately measured by the neutron imaging method (Kang et al., 2014).

The Shapiro-Wilk (SW) test statistic, *W*, was used to evaluate normality of the frequency distributions of the four BC parameters. The results of this test are given in Table 2. Asterisks (i.e. *, **, or ***) were used to denote the significance level of the SW test, and all other statistical tests employed. Table 3 lists the significance levels associated with the number of asterisks. The Quantile-Quantile (Q-Q) plots shown in Figure 4 can be used to visualize the SW results. Points on the Q-Q plot provide an indication of normality for each parameter. If the parameter is normally distributed, points will fall on a 45-degree reference line. Normality testing indicated that two of the four parameters were non-normal at the highest significance level (i.e. ***), i.e. θ_s and λ, with *W* statistics of 0.901 and 0.951, respectively. Kriging (see below) is optimal when parameters are normally distributed. Therefore, inverse (θ_s) and logarithmic (λ) transformations were used to obtain near normal distributions before moving forward with the geostatistical analyses. The success of these transformations is shown in Table 2 with *W* statistics of 0.972 and 0.993 for inverse θ_s and log λ, respectively. Although, inverse θ_s is shown to be non-normal at the lowest significance level (i.e. *) the transformed parameters Q-Q plot shown in Figure 4e demonstrates that the transformation obtained a near-normal distribution. Moving forward all trend surface and geostatistical analyses will be conducted using the transformed parameters.

Scatter diagrams, along with r, and R^2 statistics were used to indicate significant linear relationships between the BC parameters (Figure 5). In the ideal scenario, a correlation between two parameters is either +1 or -1; between these two extremes lies a spectrum of less-than-perfect relationships, including zero, which indicates no correlation (Davis, 2002). Of the six possible relationships, four were found to be significant with r and R^2 values ranging from -0.44 to 0.51 and 0.033 to 0.255, respectively. Only one of the four significant relationships was between parameters that turned out to be spatially correlated (see below), i.e. θ_s vs. ψ_a (r = -0.44, R^2 = 0.197). The implications of this relationship will be discussed further in the following sections.

4.3 Trend Identification Using the Kruskal-Wallis Test

The Kruskal-Wallis (KW) test was employed to test the equivalence of sample intervals in the x- and y-directions of the grid superimposed over the neutron images and to identify 1-dimensional trends. Quantifying trends using the KW test offers an alternative trend identification method to the trend surface analysis (which will be discussed in the next section). The results from the KW test can be found in Tables 4 and 5 for the x- and y-directions, respectively. Within each table, χ^2 was used to evaluate probability and asterisks were used to signify significance levels, with results provided for those parameters with significant trends. Table 4 shows statistical differences in the x direction for θ_s and θ_r. The trend found in θ_s was at the highest significance level (χ^2 = 36.1 ***) while the trend in θ_r was at the lowest significance level (χ^2 = 14.4 *). These results are illustrated in Figure 6, which is a plot of the interquartile range (IQR) for each sample interval in the x-direction. The plot for θ_s (Figure 6a) shows a bulge of high median values for the central

sample intervals (9.6 mm – 19.2 mm). This may be an indication that the estimation of θ_s was effected by variable path lengths caused by the cylindrical shape of the sand column. The plot for θ_r (Figure 6b) indicates a trend of increasing values across the soil column from the 3.2 mm sample interval to the 25.6 mm sample interval.

In Table 5 it is shown that ψ_a and λ experience significant differences among sample intervals in the y-direction with χ^2 values of 48.7 and 29.6 at the highest and lowest significance levels, respectively. Figure 7 shows a plot of IQR in the y-direction for the significant parameters in Table 5, this plot indicates where the statistical differences occurred for ψ_a and λ. The plot for ψ_a shown in Figure 7a shows a clear trend of increasing median values from the top to the bottom of the column. This trend may indicate that compaction occurred during sample preparation. Although a significant difference was found between sample intervals in the y-direction for λ, the plot shown in Figure 7b cannot be used to discern a specific trend direction. This may indicate that λ was unaffected by the methodology employed. These results are the first indication that the Kang et al. (2014) data collection procedure may have inadvertently caused trends in the BC parameters. To reiterate, significant differences among sample intervals were apparent in the x-direction for θ_s and in the y-direction for ψ_a, these results will become important when interpreting the kriged maps of spatially correlated parameters below.

4.4 Trend Surface Analysis and Detrending

Figure 8 shows trend surfaces for each of the BC parameters. It is evident from these initial images that the parameters have strong spatial trends. Table 6 summarizes the statistics used to determine fit and significance of the trend surfaces. The R^2 values in

Table 6 are relatively low for all of the parameters indicating poor predictive capabilities. Since the main objectives of the trend surface analysis are to determine significant trend directions and provide rough estimates of the underlying trend directions the poor R^2 values are not troublesome. The F-statistic and associated significance levels, represented by asterisks, reveal that all parameters had significant trends. Table 7 contains t-values and their associated level of significance to identify significant trend directions.

Major conclusions cannot be drawn from these results because the trend surface analysis does not account for spatial dependency and is a global interpolator, but initial inferences can be made. The θ_s trend surface shows a strong tendency of high values in the center of the sand column, possibly caused by varying thicknesses of water across the cylinder. The trend surface of ψ_a suggests trends caused by packing and/or compressive stress during sample preparation. It seems intuitive that low values would be prominent at the top of the column, whereas, compaction may lead to high values at the bottom. These results are consistent with those found during the KW test in the previous section. Both θ_r and λ exhibit similar parabolic trends that illustrate their correlation (r = 0.51), but the extremely low R^2 values 0.074 and 0.101, respectively, limit interpretation.

The descriptive statistics in Table 8 summarize the residual values for the BC parameters after trend removal. Following the detrending procedure only the fluctuations above and below the fitted surface remains, resulting in very small mean values; this is an indication that the trends were successfully removed. The trend surface analysis was conducted again on the newly created residual variable, z_{res}, for each parameter to ensure stationarity before moving forward with geostatistics. Also, to verify normality of the

residuals, the SW test was performed and the results are summarized in Table 9. The transformation and trend removal processes were successful at achieving near normal distributions for θ_s and λ. The single asterisk for θ_s indicates the parameter is still slightly positively skewed at the lowest significance level (i.e. <0.05), but to avoid employing a more complicated transformation technique the near-normal distribution was deemed acceptable.

4.5 Spatial Interpolation

4.5.1 Theoretical Model

As previously discussed there are two options for computing kriged surfaces of non-stationary data. Option one involves OK of residual data (i.e. referred to as RK in section 3.5.4), and option two entails UK of raw data. These two approaches employ different semivariograms to make predictions because the data used to compute them are not the same. This can become confusing and seem as if different semivariogram forms are used for OK as opposed to UK, but the computation of the semivariogram is the same for both methods. Tables 10 and 11 list the parameters (range, nugget, and sill) of the model with the best fit statistics (R^2 and NRMSE), and the degree of spatial dependency (DSD) defined as the ratio of the nugget to the sill. These metrics were used to evaluate the goodness of fit of the theoretical models to the empirical semivariograms and to describe the extent of spatial correlation. In this section the "best fit model" describes the model with the highest R^2 and lowest NRMSE. This signifies the optimal fit of a model to its empirical semivariogram. A true "best" semivariogram model is elusive; therefore, the most appropriate model is determined based on the statistics presented here and the following

section. The null hypothesis that there will be no spatial structure because the Kang et al. (2014) study was conducted for a nominally homogeneous material was rejected for two of the four BC parameters, θ_s and ψ_a.

In Table 10 the spherical model had the best fit statistics for all parameters, but was found to significantly fit only θ_s and ψ_a. The R^2 values for ψ_a and θ_s are 0.884 and 0.784, and NRMSE values are 0.107 and 0.127, respectively. The ranges for ψ_a (14.8 mm) and θ_s (16.2 mm) were considerably larger than those of λ (8.12) and θ_r (0 mm). According to Wei et al. (2007), the DSD is classified as strong when it is less than 0.25, moderate when it is between 0.25 and 0.75, and weak when it is greater than 0.75. As can be observed in Table 10, the DSD values for ψ_a (0.592) and θ_s (0.385) indicate moderate spatial dependency, whereas, the values for λ *(0.928)* and θ_r (1.00) indicate little to no spatial dependency.

Figure 9 shows the residual semivariograms and best fit models for each parameter. The ψ_a and θ_s display the characteristic curvilinear form of the semivariogram, while λ and θ_r are sub-linear and linear, respectively. The flat semivariograms of λ and θ_r reiterate the DSD results in Table 10. Attempting to perform spatial interpolation with these parameters would be meaningless; therefore, they are interpreted as pure nugget and will not be used to compute kriged maps. Additionally, ψ_a and θ_s were shown to be spatially correlated and will be employed in the kriging plan.

The results of Table 11 are very similar to those of Table 10 except these data seem to have a higher degree of spatial correlation. This is because when using the residual data much of the redundant (drift related) variability in the empirical semivariogram is removed (Kitanidis, 1997). When using the raw data the trend swamps the empirical

semivariogram, making it difficult to infer the true covariance function of the parameter (Kitanidis, 1997), consequently some of the trend variability is interpreted as spatial structure.

In Table 11 the spherical model was determined to have the best fit for ψ_a, θ_s, and λ; gstat was unsuccessful at fitting a model to the θ_r parameter. As seen previously, the spherical model achieved a significant fit for only ψ_a and θ_s. The R^2 and NRMSE statistics for ψ_a (0.833 and 0.121) suffered when using the raw data, but an improvement was observed for θ_s (0.859 and 0.107). A decrease in range is apparent for ψ_a (12.9 mm), but a slight increase is shown for θ_s (16.9 mm) and λ (8.12 mm). The DSD values for ψ_a (0.336), θ_s (0.425), and λ (0.859) improved when using the raw data because the underlying trend is contributing to the spatial variability of the parameters. As before, the DSD for θ_r was unity, indicating no spatial correlation. Figure 10 depicts graphs of the raw semivariograms and best fit models for the ψ_a, θ_s, and λ parameters. These semivariograms repeat the results of the residual semivariogram analyses, demonstrating that ψ_a and θ_s were spatially dependent while λ and θ_r were not.

Of the six possible relationships between the BC parameters, four were found to be significant and only one of those four was among the spatially correlated parameters, i.e. ψ_a versus θ_s. A cross-variogram was computed using the spatial relationship between ψ_a and θ_s. The residual variable for each parameter was used during the computation to deal with stationarity. Table 12 lists the parameters of the best fit model and its fit associated statistics. The spherical model was confirmed as the best fit model with an R^2 of 0.936 and NRMSE of 7.83e-2. In comparison the R^2 and NRMSE for the cross-variogram is much better

than all other semivariogram fits. The range of the cross-variogram is between the individual ranges of each parameters semivariogram. The nugget (-1.17e-3) is very small in comparison the sill (-1.69e-2) resulting in a very low DSD, 6.92e-2. The DSD for the cross-variogram indicates a much higher degree of spatial dependency than any other individual semivariogram. The nugget and sill of the cross-variogram were negative because ψ_a and θ_s were negatively correlated (r = -0.44). This indicates that as one parameter increases the other decreases; as a result, kriged maps will develop inverse features. Figure 11 shows the cross-variogram of ψ_a and θ_s with the fitted spherical model. The structural variance, the variance between the nugget and sill, is much larger for the cross-variogram than the comparable semivariograms. This higher degree of spatial structure will translate into more accurate kriged predictions later.

One of the main objectives of this study was to identify spatial structure in the BC parameters using semi/cross-variograms. Early on it was determined that any spatial structure found in the parameters must be a result of artificial heterogeneity due to measurement error or sample preparation because the Kang et al. (2014) study was conducted using a nominally homogeneous material. Considering the results of this section it is evident that the data collection procedure resulted in a moderate degree of spatial dependency for ψ_a and θ_s and a very small degree of spatial structure for λ. From these results θ_r seems to be unaffected by the methodology or the degree of spatial structure occurs on a scale smaller than the resolution of the data collection procedure.

4.5.2 Cross-validation

In an effort to determine which procedural techniques resulted in the sub-core scale heterogeneity evident in the Kang et al. (2014) data, kriged maps will be computed for the spatially dependent parameters, ψ_a and θ_s. To ensure the optimal geostatistical model is chosen to compute the final kriged map, leave-one-out cross-validation (CV) was employed. This CV is used to determine how well the kriging estimator, with a particular model, estimates values at unsampled locations. In the previous section the most appropriate semivariogram model was chosen based on its fit of the empirical data, now its predictive ability will be tested with different kriging methods. The optimal geostatistical model is a combination of the best model and kriging method. The statistical methods used to compare the CV estimated values to raw data are presented here.

The one-to-one graph of CV predicted values versus observed values was used to derive r, NRMSE, intercept (a), and slope (b) values to evaluate prediction accuracy. These statistics are given in Tables 13 and 14 for each kriging technique and its best predictive model for ψ_a and θ_s, respectively. For ψ_a (shown in Table 13) the best prediction accuracy was achieved using the spherical model for OK and CK, and the Gaussian model for UK. Of the three kriging methods, CK had the best prediction accuracy with an r of 0.726 and an NRMSE of 0.122. From Table 13, the a and b values indicate that all kriging methods are overestimating small values and underestimating large values of ψ_a. This is evident in Figure 12 by comparing the one-to-one graph of the CK method to the theoretical one-to-one line of correlation.

Table 14 shows all kriging methods achieved the best prediction accuracy using the spherical model for θ_s. As with ψ_a, CK achieved the best prediction accuracy with an r of 0.573 and an NRMSE of 0.146. Comparing the r and NRMSE values between ψ_a and θ_s it is clear that more accurate predictions were obtained for ψ_a. All kriging methods for the θ_s parameter overestimated small values and underestimated large values, as shown by a and b coefficients in Table 14. Figure 13 illustrates these results by comparing the one-to-one graph for each kriging method to the theoretical one-to-one line for θ_s.

A separate test of prediction accuracy was conducted, and the results are given in Tables 15 and 16 for ψ_a and θ_s, respectively. In these tables the "Best Prediction" column indicates the number of times the CV estimated value by a certain kriging method was closer than its competitors when estimating one of the 115 experimental data points. Percent prediction accuracy (%PA) was calculated based of the "Best Prediction" column. Table 15 shows that CK was clearly superior to the other interpolators, making the best prediction 49% of the time. These results support the data presented in Table 13 for ψ_a. As can be seen in Table 16, CK was also shown to be the best interpolator, but only by the narrowest of margins (1%). CK made the best prediction 41% of the time, whereas, UK was only slightly less accurate with 40%. The results presented here support the results given in Table 14 for θ_s.

From the information provided thus far the optimal geostatistical model consists of both the spherical model and CK form for ψ_a and θ_s. As stated by Goovaerts (1997) there is no "best" semivariogram model; as a result, the user should determine the model based on the information available and the objective to be accomplished. In this study the most

appropriate semivariogram model and kriging form were chosen based on the information put forth in the previous section and the CV results provided here. Now that the optimal geostatistical model (both semivariogram model and kriging form) has been chosen, kriged maps of the spatially dependent parameters, ψ_a and θ_s, can be computed with confidence.

4.5.3 Kriging

Up to this point speculations have been made as to the principal causes of spatial dependency in ψ_a and θ_s. Now an in depth look at the spatial structure of the parameters can be made using the final kriged maps given in Figure 14. Starting with ψ_a, a layer of low values is shown at the top of the column. Possible causes of this include packing and compressive stress imposed by the weight of the overlying material in the sand column. During the packing process sand is incrementally added to the column in order to minimize its impact on fluid movement. Inevitably unnatural alterations are made to the sand pack which can cause method-based heterogeneity in a nominally homogeneous material. With each addition of sand, larger particles have the potential to move to the top resulting in a layer of low ψ_a values. In the kriged map of ψ_a this is only shown to have occurred at the top of the column. Otherwise layering would be evident throughout the sand column. Also, in general material will be looser and less compact at the top of the column because the grains are being subjected to less compressive stress. A second feature in the kriged map of ψ_a that is obvious is a large region of high values at the bottom of the column. Compaction of sand due to compressive stress is an intuitive explanation for these high values. As more material is added to the column during packing the weight of the overlying sand increases causing compaction. Greater compaction results

in larger capillary pressures needed to drain pores. The features found in the kriged map of ψ_a are a result of inconsistency in core packing; therefore, the alternative hypothesis can be accepted for this parameter.

The most obvious feature in the kriged map of ψ_a is the zone of low values in its center. A physical process cannot be used to explain this observation; therefore, the relationship between ψ_a and θ_s is used to propose an answer. The scatter diagram comparing ψ_a and θ_s (Figure 5a) revealed a significant negative (r = -0.44 **) correlation between these parameters. The ψ_a value represents the pressure at which the largest pore throats begin to drain. When compaction occurs both the porosity and the size of the largest pore throats are affected. During compaction porosity decreases and the largest pore throats disappear preferentially. A decrease in porosity results in a reduction in the size of the largest pore throats and a corresponding increase in the ψ_a value; therefore, ψ_a is inversely related to porosity. Since θ_s is essentially a measure of porosity, an explanation of the features in the kriged map of θ_s will provide an answer for the concentration of low ψ_a values in the center of its kriged map.

Looking at the kriged map for θ_s, a central tendency of high values is evident, with a hot spot of large values in the center of the column (Figure 14b). This is the inverse of the kriged map of ψ_a (Figure 14a). The central tendency of high values mimics the cylindrical shape of the aluminum column used to contain the sand pack. The shape of the trend suggests that the image analysis process may have contributed to the spatial structure in θ_s. Before point values of volumetric water content were calculated and paired with capillary pressures to compute water retention curves, a normalization process was performed on the raw neutron radiographs. This normalization process was done to

remove the variable path lengths associated with the cylindrical column shape. The fact that a central tendency is still apparent in the kriged map of θ_s suggests that this normalization procedure was not 100% in removing the effects of different path lengths.

To understand the sub-core scale heterogeneity shown in the kriged map of θ_s a detailed explanation of the imaging process is given here. Wet neutron radiographs were acquired during the drainage process of the sand column. After the final drainage step, the sand column was oven drained and imaged again to obtain a dry image of the sand pack. The collected raw images (wet and dry) were normalized with respect to reference images of the open beam and dark field to correct for background noises, inhomogeneities in the beam detector, and fluctuations in the neutron flux (Kang et al., 2014). Normalized wet images were then divided by normalized oven dry images to calculate water thicknesses across the sand column. The water thickness of each pixel was computed using an attenuation and correction coefficient previously obtained at the HFIR CG 1-D. Subsequently, volumetric water content was calculated for individual pixels by dividing the measured water thickness by the beam path length, which is the chord length for a circle (Kang et al, 2014). This normalization process essentially flattens out the sand column. The central tendency of high values in the kriged map of θ_s illustrate that the normalization process was not completely effective.

The hotspot of high values in the center of the kriged map for θ_s can also be partly attributed to neutron scattering. Neutrons are attenuated by the water in the sand column resulting in increased scatter and decreased neutron intensity. Raw neutron radiographs are a measure of neutron intensity; therefore, little or no detection of neutrons

corresponds to high water content. Water thickness is greatest in the center of the column because this is its widest point. The water content calculated for a pixel in the center of the column can be the same as the water content calculated for a pixel on the edge with differing amounts of water by virtue of the cylindrical shape of the column. This is why the normalization process is necessary. As previously stated, with more water increased scattering will occur. Since larger quantities of water are present in the center of the column more scattering takes place. The increased scatter in the center of the column causes a decrease in detection of neutron intensity. This translates to higher water contents even if the water content in the center is the same as along the edge. In effect the varying thickness of water along the column results in both the central tendency and hot spot of high values in the kriged map of θ_s.

4.6 Trend Correction and Validation

Tables 17 through 25 summarize the results of steps 1 through 8 listed in section 3.7. A summary of these results is given because they are very similar to the results presented in sections 4.41 and 4.4.2 for the Kang et al. (2014) data set. The results of Tables 17 through 25 are specific to the data set created during the CV procedure for ψ_a and θ_s. These results are given to describe the CV data set, created by the optimal geostatistical model, used to aid in correction of non-geologic spatial structure within ψ_a and θ_s. First, the results of the trend correction process will be presented. Then, validation results will be given to determine if the correction process was successful.

Table 17 gives the descriptive statistics of the CV data set, and Table 18 provides the results of normality testing. As with the Kang data, ψ_a is normally distributed and θ_s is not.

An inverse transformation was applied to θ_s before moving forward with trend surface analysis. The results of this transformation are given in Table 18; normality was achieved with a *W* statistic of 0.982. Table 19 shows that significant trends were evident for both parameters. Significant trend directions for ψ_a include x, x*y, x^2, and y^2; whereas, the significant trends for θ_s were limited to x and x^2 as shown by Table 20. Classical descriptive statistics are given in Table 21 for the trend surface residuals computed for the CV data set. Normal and near normal distributions are shown for ψ_a and θ_s, respectively in Table 22. The spherical model was fit to the cross-variogram of the CV data; these results are presented in Table 23. In comparison to Table 12 a slightly worse fit of the spherical model to the cross-variogram was indicated by an R^2 and NRMSE, of 0.897 and 0.109, respectively. Also, the DSD was slightly higher for the CV data than the Kang data. For consistency the optimal geostatistical model determined in section 4.4.2 was also used here to perform CV and compute kriged maps of both parameters.

CV was used to assess the predictive ability of the spherical model and CK form for the CV data. This was mostly a precautionary step to ensure the quality of predictions made for the CV data set. Tables 24 and 25 summarize the cross-validation results for the CV data set of ψ_a and θ_s, respectively. The data presented in these tables indicate better prediction accuracy for both ψ_a and θ_s. With r values of 0.824 and 0.716 the predictive ability using the CV data set was shown to dramatically improve for ψ_a and θ_s, respectively. The NRMSE results reiterate the improved predictive ability with values of 0.102 and 0.126 for ψ_a and θ_s. As before, a and b coefficients for both parameters indicate over predication of small values and under prediction of large values. Figure 15 shows kriged maps of the CV data set for ψ_a and θ_s. As expected, the maps were very similar to those in Figure 14.

The sub-core scale heterogeneity found in ψ_a and θ_s was shown to be caused by methods based processes in section 4.4.3. To remove those trends and discover the actual variability, a map of the detrended variable, z_{ra} (created in section 3.7), was computed. Since the material used in the Kang et al. (2014) study was homogeneous, no spatial structure was anticipated after trend correction in z_{ra}. Semivariograms were used to test for spatial dependency in the z_{ra} variable for ψ_a and θ_s. The semivariograms computed for z_{ra} were horizontal and contained no structural variance; therefore, they were determined to be pure nugget and confirm the initial speculation. Classical descriptive statistics for the z_{ra} variable is given in Table 26 for both ψ_a and θ_s.

Because kriging is an exact interpolator (i.e. the predicted and observed values are equal at all measurement locations) dependent upon spatial structure, geostatistics cannot be employed to produce maps of the z_{ra} variable; therefore, the map shown in Figure 16 was created using the residual values between the kriged maps computed for the Kang and CV data sets. Figure 16 is a result of first computing the average of all kriged predictions for the original Kang dataset. Then adding it to the difference found by subtracting the kriged values calculated using the Kang data set and the CV data set. As a result, the z_{ra} values for both ψ_a and θ_s can be found at measured locations on the corrected map in Figure 16. From this map it is evident there is no spatial dependency in either parameter once the spatial (non-geologic trends) are removed. The variability shown in the map for each parameter is relatively consistent and displays no correlation between values as a result of distance.

In order to validate the trend correction process, the z_{ra} was compared to an independent dataset collected by Cropper et al. (2014). In Tables 27 and 28 z_{ra} is referred

to as "Krige". Table 27 gives the Mann-Whitney (MW) W statistic and corresponding significance level denoted by asterisks (Table 2). Three data sets are compared to the Cropper data, two of which consist of all BC parameters, the Kang and Trend data, and the Krige data which contains only the spatially dependent parameters ψ_a and θ_s. Since trend surface analysis is as an inexact interpolator the Trend data set was generated during the trend identification and removal process. The Cropper data set contains nine values while each of the three comparison sets consists of 115 data points. As can be seen in Table 27, median values of the ψ_a, θ_s, and θ_r parameters were significantly different from the median values of the Cropper parameters. This result indicates that, in terms of median values, the data collection process has unintentionally effected the computation of three BC parameters (ψ_a, θ_s, and θ_r) and that the trend correction process used in this study to remove non-geologic spatial dependency was not successful. Figure 17 is an illustration of these results; it compares the IQR of each BC parameter and its corresponding data set. From the figure it is apparent that the upper and lower quartiles for ψ_a, θ_s, and θ_r do not overlap the IQR of the Cropper data. Table 28 presents the Kolmogorov-Smirnov (KS) two sample test D statistic and significance levels. These results confirm the conclusions drawn from Table 27. In this case, the frequency distributions for ψ_a, θ_s, and θ_r were significantly different from the frequency distributions for these parameters from the Cropper data set, while the frequency distribution for λ was not.

CHAPTER 5
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Precise characterization of sub-core scale heterogeneity is essential for accurate prediction of fluid movement in the subsurface. This research focused on improving our ability to quantify sub-core scale features and on identifying their spatial structure using a combination of neutron imaging and geostatistics. There are two major conclusions to be drawn from this study. First, geostatistics was successfully applied at the sub-core scale to characterize the spatial variability in point Brooks and Corey (BC) soil water retention parameters obtained by Kang et al. (2014) using neutron radiography. Second, the spatial dependency found in the BC parameters was likely attributed to sample preparation and image analysis, rather than to the hydrogeological phenomena. Ultimately this study offers a way to visualize areas of potential improvement for the Kang et al. (2014) data collection procedure. This study also proposes a technique for correcting methods based spatial dependency. The goal of the correction process was to remove the effects of methodology from each spatially dependent BC parameter and expose the actual variability present within the parameter. In doing this the true controls on flow and transport at the sub-core scale can be identified. The main results used to arrive at these conclusions are discussed below along with some suggestions for future work.

Histograms illustrated substantial variability within the BC parameters for this nominally homogeneous material. Kang et al. (2014) offered the packing procedure as explanation for the sub-core scale variability. In this study geostatistics was employed to determine if the sub-core scale variability evident in parameter histograms is spatially

correlated. This is an important task because spatially similar regions, whether at the sub-core scale or field scale, will dictate subsurface flow.

The first steps in the geostatistical analysis were to determine linear relationships between parameters and test for normality. Four of the six possible relationships between parameters were found to be significant, but later it was determined only one relationship was between spatially correlated parameters, i.e. θ_s vs. ψ_a ($r = -0.44$, $R^2 = 0.197$). The Shapiro-Wilk (SW) test indicated that both distributions were significantly different from a normal distribution. As a result these parameters were transformed to yield normal and nearly-normal distributions before moving forward with the geostatistical analysis.

To satisfy the intrinsic hypothesis and ensure stationarity a trend surface analysis was conducted. The trend surface analysis showed that all four BC parameters contained significant trends. The trend surfaces (Figure 8) provided some of the first evidence that the sample packing and data collection procedures had impacted the BC parameters. The trends were effectively removed by following the detrending procedure suggested by Vieira et al. (2010). Moving forward the residuals computed during the detrending process were used in place of the original Kang data for ordinary kriging (OK) and co-kriging (CK).

The most critical part of conducting a geostatistical analysis is calculating the empirical semivariogram and fitting it with a theoretical model. The empirical semivariogram identifies spatial correlation by plotting the semivariance versus distance. If this relationship results in a horizontal line no spatial correlation is evident in the parameter. This was true for two of the four parameters, i.e. θ_r and λ. This lack of spatial correlation indicates that either these parameters were unaffected by the sample

preparation/measurement method or the variability within these parameters exists at a scale smaller than the sampling distance. In contrast, spatial dependency was clearly evident in the ψ_a and θ_s parameters from their semivariograms; therefore, the null hypothesis was rejected. The spherical model achieved the best fit to the empirical semivariogram and cross-variogram for ψ_a and θ_s based on the R^2 and NRMSE statistics. It was shown by the ratio of the nugget to the sill that ψ_a possessed a higher degree of spatial dependency than θ_s. The first objective of this research listed in section 1.2 was to provide semivariograms of BC parameters; this was accomplished for ψ_a and θ_s.

Leave-one-out cross-validation (CV) was used to determine an optimal geostatistical model to compute the final kriged maps of ψ_a and θ_s. The geostatistical model consists of the most appropriate semivariogram model and kriging form. Several statistical metrics were used to determine the optimal geostatistical model. In the end, using the spherical model and CK form resulted in the best prediction accuracy. CV was also used to generate a data set that would later be used to correct for non-geologic spatial structure.

The final CK maps of ψ_a and θ_s can be seen in Figure 14. They were computed based on the significant linear relationship found between these two parameters. The initial suspicion put forth by Kang et al. (2014) regarding the cause of sub-core scale variability was confirmed. The effects of packing and compressive stress were evident in the CK map for ψ_a; therefore, the alternate hypothesis was accepted. Low values were apparent at the top of the column, while generally higher values occurred at the bottom. Packing and compressive stress could be used to explain the majority of trends in the map. These findings are consistent with the work of Lewis and Sjöstrom, 2010, who stated that packing

of homogeneous material can cause spatial heterogeneity in the form of preferential flow paths which can effect transport of solutes through a porous medium. The negative correlation between ψ_a and θ_s was used to explain the concentration of low values in the center of the ψ_a kriged map. These low values corresponded to a hot spot of high values in the center of the θ_s kriged map. The trends found in this map were explained as a result of the image analysis procedure and could not be attributed to a physical procedure such as packing. Both the hot spot and central tendency of high values in the kriged map of θ_s were likely caused by varying thickness of water across the sand column due to its cylindrical shape. Because of the shape of the sand column, a volumetric water content computed for a single pixel in the center of the column could contain a larger quantity of water than a pixel on the edge with the same volumetric water content. The kriged map of θ_s shows the image processing technique used to eliminate this effect was not completely successful. Also, greater water thicknesses in the center of the column are thought to be the cause of the hot spot of high values in the center of the column. Because neutrons are attenuated by the hydrogen in water more scattering occurs with an increased presence of water.

The next logical step in this research was to attempt to remove the spatial structure caused by methodology found in ψ_a and θ_s. In order to accomplish this, a detrending technique similar to the one used to achieve stationarity was employed. During the trend surface analysis computing residual values was simple because it is an inexact interpolation method. Because kriging is an exact interpolator a second set of data generated using leave-one-out CV was used to compute residual values by finding the difference between the original Kang data and the CV data set. The average of the observed data was added to the residuals to create a new variable z_{ra}. This variable was mapped for

both parameters and is shown in Figure 16. It represents the true geologic variability in ψ_a and θ_s once the trends caused by methodology were removed. The maps of ψ_a and θ_s after trend correction show no spatial correlation. To verify this, semivariograms of the z_{ra} variable for both parameters resulted in pure nugget effects. To validate the correction process the z_{ra} variable computed for ψ_a and θ_s was compared to an independent data set of the 9 values for each BC parameter collected by Cropper (2014).

The Mann-Whitney (MW) and Kolmogorov two sample (KS) statistics were used to test the null hypothesis that the corrected and independent data had the same median values and frequency distributions, respectively. The MW test is more sensitive to changes in median values; whereas, the KS test has more power to determine changes in the shape of the distribution. Both test statistics ultimately indicated a significant difference between the two group's distributions. From these results it can be deduced that the trend correction process was unsuccessful.

This research successfully identified spatial dependency in point BC parameters and in doing so determined possible experimental causes of this sub-core scale heterogeneity in a nominally homogenous sand material. Also, spatial structure was found to be caused by inconsistency in packing for ψ_a. Although the attempted correction process was unsuccessful, new information was gained about the data collection procedure conducted by Kang et al. (2014). The newly acquired information from this research can be used to improve upon the work performed by Kang et al. (2014) before the technique is extended to other natural materials. Identifying the effects of methodology is necessary to improve the modeling of sub-core scale heterogeneity resulting from geology in future studies.

Future work should include a revised correction process. Geostatistics is a viable option for overcoming the effects of methodology in the BC parameters. In the future an improved means of discriminating between methods based and actual spatial structure is needed. Now that the effects of varying water thickness have been show in θ_s, for future work a new geostatistical weighting scheme could be developed to combat it. For instance, if it is assumed that the central tendency of high values is in the center of the column was caused by its cylindrical shape; it might be possible to assign increasingly larger weights to interpolated values from the center to the edge for θ_s based on the formula for a chord of a circle. A similar weighting scheme could also be applied to ψ_a to remove the trend caused by compressive stress as a linear function of depth. Weights would then be applied to ψ_a in the opposite direction of these trends. This idea offers a quantitative solution for the existing trends in the BC parameters.

Possible physical solutions for future exploration may be to employ consolidated materials and/or methods to remove the effects of repacking. Such methods may include using low disturbance techniques that retain sedimentary layers, for example employing push type coring and frozen cores. The use of rectangular sample geometry could be used to eliminate variable neutron pathways. Because employing a rectangular column can introduce edge effects, the column would be large enough so that the imaging procedure could be focused on the center of the column (i.e. the unaffected region). Additionally, the effectiveness of this correction method could be tested by preparing sand columns with deliberate heterogeneity (e.g. layering), and then attempting to determine spatial variability within BC parameters using geostatistics and neutron imaging.

LIST OF REFERENCES

Aylmore, L.A.G., and V. Rasiah. 1998. Computed tomography data on soil structural and hydraulic parameters assessed for spatial continuity by semivariance geostatistics. Aust. J. Soil Res. 36(3): 485+.

Bohling, G. 2005. Introduction to Geostatistics and Variogram Analysis. Kans. Geol. Surv.: 20.

Bromly, M., C. Hinz, and L. a. G. Aylmore. 2007. Relation of dispersivity to properties of homogeneous saturated repacked soil columns. Eur. J. Soil Sci. 58(1): 293–301.

Brooks, R.H., and A.T. Corey. 1964. Hydraulic Properties of Porous Media. Colorado State University.

Carducci, C.E., G. César de Oliveira, N. Curi, R.J. Heck, and D.F. Rossoni. 2014. Scaling of pores in 3D images of Latosols (Oxisols) with contrasting mineralogy under a conservation management system. Soil Res. 52(3): 231–243.

Cheng, C.L., M. Kang, E. Perfect, S. Voisin, J. Horita, H.Z. Bilheux, J.M. Warren, D.L. Jacobson, and D.S. Hussey. 2012. Average Soil Water Retention Curves Measured by Neutron Radiography. Soil Sci. Soc. Am. J. 76(4): 1184–1191.

Cressie, N.A.C. 1993. Statistics for Spatial Data. Wiley, New York.

Cropper, S. 2014. Comparison of Point and Average Capillary Pressure - Saturaion Functions for Porous Media. Dr. Diss.Available at http://trace.tennessee.edu/utk_graddiss/3115.

Davis, J.C. 2002. Statistics and Data Analysis in Geology. 3rd ed. John Wiley & Sons, New York.

Deinert, M.R., J.-Y. Parlange, T. Steenhuis, J. Throop, K. Ünlü, and K.B. Cady. 2004. Measurement of Fluid Contents and Wetting Front Profiles by Real-Time Neutron Radiography. J. Hydrol. 290(3–4): 192–201.

Dubrule, O. 2003. Geostatistics for Seismic Data Integration in Earth Models. Society Of Exploration Geophysicists, Tulsa, OK.

Duijn, C.J.V., J. Molenaar, and M.J.D. Neef. 1995. The effect of capillary forces on immiscible two-phase flow in heterogeneous porous media. Transp. Porous Media 21(1): 71–93.

Dunlap, L.E., and J.M. Spinazola. 1984. Interpolating Water-Table Altitudes in West-Central Kansas Using Kriging Techniques. Geol. Surv. Water-Supply Pap. USA No 2238Available at http://agris.fao.org/agris-search/search.do?f=2012/OV/OV201204180004180.xml;US19850071841 (verified 2 April 2014).

Eaton, T.T. 2006. On the importance of geological heterogeneity for flow simulation. Sediment. Geol. 184(3–4): 187–201.

Ecker, M.D. 2003. Geostatistics: past, present and future. Environmetrics Dev. Auspices UNESCO Encycl. Life Support Syst. EOLSS Ed AH El-Shaarawi J Jureckova Eolss Publ. Oxf. UK Httpwww Eolss Net.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford, NY.

Goovaerts, P. 1998. Ordinary Cokriging Revisited. Math. Geol. 30(1): 21–42.

Grevers, M.C., and E. De Jong. 1994. Evaluation of soil-pore continuity using geostatistical analysis on macroporosity in serial sections obtained by computed tomography scanning. Tomogr. Soil-Water-Root Process. (tomographyofsoi): 73–86.

De Gryze, S., L. Jassogne, J. Six, H. Bossuyt, M. Wevers, and R. Merckx. 2006. Pore Structure Changes During Decomposition of Fresh Residue: X-Ray Tomography Analyses. Geoderma 134(1–2): 82–96.

Harned, D.A. 1995. Effects of agricultural land-management practices on water quality in northeastern Guilford County, North Carolina, 1985-90. U.S. Dept. of the Interior, U.S. Geological Survey ; U.S. G.P.O. ; Copies can be purchased from U.S. Geological Survey, Information Services,.

Hartkanmp, D., K. De Beurs, A. Stein, and J. White. 1999. Interpolation Techniques for Climate Variables. Nat. Resour. Group.

Houston, A.N., W. Otten, P.C. Baveye, and S. Hapca. 2013. Adaptive-window indicator kriging: A thresholding method for computed tomography images of porous media. Comput. Geosci. 54: 239–248.

H. Saito, K. Seki, and J. Šimůnek. 2009. An alternative deterministic method for the spatial interpolation of water retention parameters. Hydrol. Earth Syst. Sci. 13(4): 453.

Huysmans, M., and A. Dassargues. 2006. Stochastic analysis of the effect of spatial variability of diffusion parameters on radionuclide transport in a low permeability clay layer. Hydrogeol. J. 14(7): 1094–1106.

Iqbal, J., J.A. Thomasson, J.N. Jenkins, P.R. Owens, and F.D. Whisler. 2005. Spatial Variability Analysis of Soil Physical Properties of Alluvial Soils. Soil Sci. Soc. Am. J. 69(4): 1338.

Isaaks, E.H., and R.M. Srivastava. 1989. Applied Geostatistics. Oxford University Press, New York.

Jalbert, M., J. Dane, and J. Liu. 1999. TrueCell: Physical point Brooks–Corey parameters using pressure cell data. Users guide for version 1.2. Spec Rep Dep Agron Soils Auburn Univ Auburn AL.

Journel, A.G., and C.J. Huijbregts. 1978. Mining Geostatistics. Academic Press.

Kang, M., E. Perfect, C.L. Cheng, H.Z. Bilheux, J. Lee, J. Horita, and J.M. Warren. 2014. Multiple Pixel-Scale Soil Water Retention Curves Quantified by Neutron Radiography. Adv. Water Resour. 65: 1–8.

Kitanidis, P.K. 1997. Introduction to Geostatistics: Applications in Hydrogeology. Cambridge University Press.

Krause, M., S. Krevor, and S.M. Benson. 2013. A Procedure for the Accurate Determination of Sub-Core Scale Permeability Distributions with Error Quantification. Transp. Porous Media 98(3): 565–588.

Kumar, V., and D. Remadevi. 2006. Kriging of Groundwater Levels – A Case Study. J. Spat. Hydrol. 6(1): 81–94.

Lakhankar, T., A.S. Jones, C.L. Combs, M. Sengupta, T.H. Vonder Haar, and R. Khanbilvardi. 2010. Analysis of Large Scale Spatial Variability of Soil Moisture Using a Geostatistical Method. Sensors 10(1): 913–932.

Lehmann, E.L., and H.J.M. D'Abrera. 2006. Nonparametrics: Statistical Methods Based on Ranks. 1st ed. 1975. Revised edition 2006 edition. Springer, New York.

Lewis, J., and J. Sjöstrom. 2010. Optimizing the experimental design of soil columns in saturated and unsaturated transport experiments. J. Contam. Hydrol. 115(1-4): 1–13.

Li, B., and S.M. Benson. 2014. Small-scale Heterogeneities and Buoyancy-driven CO2 Migration in Geological Storage. Energy Procedia 63: 3608–3615.

Li, B., H.A. Tchelepi, and S.M. Benson. 2013. Influence of capillary-pressure models on CO2 solubility trapping. Adv. Water Resour. 62, Part C: 488–498.

Lloyd, C.D. 2014. Exploring Spatial Scale in Geography. John Wiley & Sons, New York.

Majani, B.S. 2007. Analysis of External Drift Kriging Algorithm with application to precipitation estimation in complex orography. Int. Inst. Geo-Inf. Sci. Earth Obs. Enschede Neth.

Materon, G. 1971. The Theory of Regionalized Variables and Its Applications.

Matheron, G. 1963. Principles of geostatistics. Econ. Geol. 58(8): 1246–1266.

Murata, S., and T. Saito. 2003. Environmental Rock Engineering: Proceedings of the First Kyoto International Symposium on Underground Environment, Kyoto, Japan, 17-18 March 2003. CRC Press.

Neiman, F. 2007. Spatial Data Analysis in Archaeology. Available at http://people.virginia.edu/~fn9r/anth589b/index.html.

Oh, W., and W.B. Lindquist. 1999. Image Thresholding by Indicator Kriging. IEEE Trans. Pattern Anal. Mach. Intell. 21(7): 590–602.

Olea, R.A. 1991. Geostatistical Glossary and Multilingual Dictionary. Oxford University Press.

Oliver, M. 2010. Geostatistical Applications for Precision Agriculture. 2010th ed. Springer, Dordrecht; New York.

Oliviera, I.B., A.H. Demond, and A. Salehzadeh. 1996. Packing of Sands for the Production of Homogeneous Porous Media. Soil Sci. Soc. Am. J. 60(1): 49.

Orjuela-Matta, H.M., Y. Rubiano Sanabria, and J.H. Camacho-Tamayo. 2012. Análisis Espacial de la Infiltración en un Oxisol de los Llanos Orientales de Colombia. Chil. J. Agric. Res. 72(3): 404–410.

Papafotiou, A., R. Helmig, J. Schaap, P. Lehmann, A. Kaestner, H. Flühler, I. Neuweiler, R. Hassanein, B. Ahrenholz, J. Tölke, A. Peters, and W. Durner. 2008. From the Pore Scale to the Lab Scale: 3-D Lab Experiment and Numerical Simulation of Drainage in Heterogeneous Porous Media. Adv. Water Resour. 31(9): 1253–1268.

Papritz, A., H.R. Künsch, and R. Webster. 1993. On the Pseudo Cross-Variogram. Math. Geol. 25(8): 1015–1026.

Pebesma, E.J. 2004. Multivariable Geostatistics in S: The Gstat Package. Comput. Geosci. 30(7): 683–691.

Perfect, E., C.-L. Cheng, M. Kang, H.Z. Bilheux, J.M. Lamanna, M.J. Gragg, and D.M. Wright. 2014. Neutron Imaging of Hydrogen-Rich Fluids in Geomaterials and Engineered Porous Media: A Review. Earth-Sci. Rev. 129: 120–135.

Perret, J., S.O. Prasher, A. Kantzas, and C. Langford. 1998. Characterization of Macropore Morphology in a Sandy Loam Soil Using X-Ray Computer Assisted Tomography and Geostatistical Analysis. Can. Water Resour. J. Rev. Can. Ressour. Hydr. 23(2): 143–165.

Perrin, J.-C., and S. Benson. 2009. An Experimental Study on the Influence of Sub-Core Scale Heterogeneities on CO2 Distribution in Reservoir Rocks. Transp. Porous Media 82(1): 93–109.

Peters, E.J., and N. Afzal. 1992. Characterization of Heterogeneities in Permeable Media with Computed Tomography Imaging. J. Pet. Sci. Eng. 7(3–4): 283–296.

Pini, R., and S.M. Benson. 2013. Characterization and scaling of mesoscale heterogeneities in sandstones. Geophys. Res. Lett. 40(15): 3903–3908.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R. Pini, S.C.M.K. 2012. Capillary heterogeneity in sandstones rocks during CO2/water core-flooding experiments. Energy Procedia 37: 11375–.

SAS Institute, Inc. 1996. SAS/STAT Technical Report: Spatial Prediction Using the SAS System. SAS Institute, Inc., Cary, NC.

Shapiro, S., and M. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52(3/4): 591–611.

Tang, H. 2005. Geostatistical Integration of Geophysical, Well Bore and Outcrop Data for Flow Modeling of a Deltaic Reservoir Analogue. Available at http://etd.lsu.edu/docs/available/etd-07032005-090208/ (verified 13 April 2015).

Tumlinson, L.G., H. Liu, W.K. Silk, and J.W. Hopmans. 2008. Thermal Neutron Computed Tomography of Soil Water and Plant Roots. Soil Sci. Soc. Am. J. 72(5): 1234.

Vasin, M., P. Lehmann, A. Kaestner, R. Hassanein, W. Nowak, R. Helmig, and I. Neuweiler. 2008. Drainage in Heterogeneous Sand Columns with Different Geometric Structures. Adv. Water Resour. 31(9): 1205–1220.

Vieira, S.R., J.R.P. de Carvalho, M.B. Ceddia, and A.P. González. 2010. Detrending non stationary data for geostatistical applications. Bragantia 69: 01–08.

Vieira, S.R., J.R.P. Carvalho, and A.P. González. 2009. Jack knifing for semivariogram validation. Bragantia 69: 97–105.

Vieira, S.R., J.L. Hatfield, D.R. Nielsen, and J.W. Biggar. 1983. Geostatistical theory and application to variability of some agronomical properties. Hilgardia 51(3): 1–75.

Warrick, A., D. Myers, and D. Nielsen. 1986. Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods. 2 edition. American Society of Agronomy-Soil Science Society of America, Madison, WI.

Zinn, B., and C.F. Harvey. 2003. When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. Water Resour. Res. 39(3): 1051.

APPENDICES

Appendix 1. Tables

Table 1: Descriptive statistics of observed BC parameters (Equation 1)

Parameter	Unit	N	Mean	Median	Standard Deviation	Coefficient of Variation
ψ_a	cm	115	17.5	17.5	0.820	4.69e-2
θ_s	cm^3/cm^3	115	0.374	0.464	4.44e-2	0.119
λ	-	115	7.74	7.42	2.50	0.323
θ_r	cm^3/cm^3	115	2.66e-2	2.55e-2	1.08e-2	0.405

Table 2: Asterisks are used to denote the level of significance for multiple statistical tests presented in the study; for example, *** indicates significance at the 99.9% confidence level, ** indicates significance at the 99% confidence level, * indicates significance at the 95% confidence level, and non-significant (confidence levels < 95%) results are indicated using NS

Significance Levels	Asterisks
< 0.001	***
< 0.01	**
< 0.05	*
Not Significant	NS

Table 3: SW *W* statistic used to test if the BC parameters come from a normal distribution before employing trend surface analysis and moving forward with the kriging analyses

Parameter	Unit	N	W
ψ_a	cm	115	0.984
θ_s	cm^3/cm^3	115	0.901 ***
λ	-	115	0.951 ***
θ_r	cm^3/cm^3	115	0.983
$1/\theta_s$	cm^3/cm^3	115	0.972 *
$Log_{10}(\lambda)$	-	115	0.993

Table 4: KW results for sample intervals in the x-direction

Parameter	Unit	χ^2
ψ_a	cm	NS
θ_s	cm^3/cm^3	36.1 ***
λ	-	NS
θ_r	cm^3/cm^3	14.4 *

Table 5: KW results for sample intervals in the y-direction

Parameter	Unit	χ^2
ψ_a	cm	48.7 ***
θ_s	cm^3/cm^3	NS
λ	-	29.6 *
θ_r	cm^3/cm^3	NS

Table 6: Trend surface analysis results used to identify significant trends in the BC parameters before kriging

Parameter	Unit	N	R^2	F-statistic
ψ_a	cm	115	0.342	14.3 ***
θ_s	cm^3/cm^3	115	0.274	21.2 ***
λ	-	115	0.121	5.11 ***
θ_r	cm^3/cm^3	115	0.074	4.45 **

Table 7: t-values listed with their corresponding significance levels to determine specific trend directions

Parameter	Unit	N	x	y	x*y	x^2	y^2
ψ_a	cm	115	-3.93 ***	NS	4.41 ***	2.30 *	-6.62 ***
θ_s	cm^3/cm^3	115	5.90 ***	NS	NS	-5.77 ***	NS
λ	-	115	NS	NS	3.36 **	-2.61 *	-3.86 ***
θ_r	cm^3/cm^3	115	NS	NS	2.93 **	NS	-2.30 *

Table 8: Descriptive statistics of trend surface residuals, very small mean values signify successful removal of trends within the BC parameters

Parameter	Unit	N	Mean	Median	Standard Deviation
ψ_a	cm	115	-1.74e-11	3.67e-2	0.665
θ_s	cm^3/cm^3	115	4.35e-11	-6.93e-3	3.88e-2
λ	-	115	-6.09e-11	-7.32e-4	0.297
θ_r	cm^3/cm^3	115	-2.61e-11	-7.98e-4	1.04e-2

Table 9: SW *W* statistic and significance level for trend surface residuals before kriging, normal to near normal distributions of BC parameters before kriging will result in optimal interpolation

Parameter	Unit	N	W
ψ_a	cm	115	0.987
θ_s	cm^3/cm^3	115	0.976 *
λ	-	115	0.989
θ_r	cm^3/cm^3	115	0.980

Table 10: Semivariogram model results computed for residual BC parameters

Parameter	Unit	Model	Nugget $(unit)^2$	Sill $(unit)^2$	Range (mm)	R^2	NRMSE	DSD
ψ_a	cm	Spherical	0.206	0.535	14.8	0.884 ***	0.107	0.385
θ_s	cm^3/cm^3	Spherical	1.09e-3	1.84e-3	16.2	0.784 ***	0.127	0.592
λ	-	Spherical	4.78	5.63	8.12	0.119	0.268	0.928
θ_r	cm^3/cm^3	Spherical	1.15e-4	1.15e-4	NA	2.49e-2	0.316	1.00

Table 11: Semivariogram model results computed for observed BC parameters

Parameter	Unit	Model	Nugget $(unit)^2$	Sill $(unit)^2$	Range (mm)	R^2	NRMSE	DSD
ψ_a	cm	Spherical	0.186	0.553	12.9	0.833 ***	0.121	0.336
θ_s	cm^3/cm^3	Spherical	4.15e-2	9.77e-2	16.9	0.859 ***	0.107	0.425
λ	-	Spherical	1.47e-2	1.71e-2	9.67	0.188	0.266	0.859
θ_r	cm^3/cm^3	-	1.14e-4	1.14e-4	NA	NA	NA	1.00

Table 12: Cross-variogram with spherical model computed for the residual values of ψ_a versus θ_s

Parameter	Unit	Model	Nugget	Sill	Range (mm)	R^2	NRMSE	DSD
ψ_a vs. θ_s	-	Spherical	-1.17e-3	-1.69e-2	15.5	0.936 ***	7.83e-2	6.92e-2

Table 13: CV results for ψ_a – compares different kriging methods and their semivariogram model which resulted in the best prediction accuracy as indicated by r and NRMSE

Kriging Method	Model	r	NRMSE	a	b
OK	Spherical	0.704	0.126	8.47	0.515
UK	Gaussian	0.681	0.130	9.85	0.436
CK	Spherical	0.726	0.122	7.87	0.550

Table 14: CV results for θ_s – compares different kriging methods and their semivariogram model which resulted in the best prediction accuracy as indicated by r and NRMSE

Kriging Method	Model	r	NRMSE	a	b
OK	Spherical	0.541	0.150	0.257	0.310
UK	Spherical	0.503	0.155	0.278	0.246
CK	Spherical	0.573	0.146	0.241	0.356

Table 15: Presents the results of the kriging form with the best prediction accuracy for ψ_a

Kriging Method	Best Prediction	%PA
OK	32/115	28
UK	27/115	23
CK	56/115	49

Table 16: Presents the results of the kriging form with the best prediction accuracy for θ_s

Kriging Method	Best Prediction	%PA
OK	22/115	19
UK	46/115	40
CK	47/115	41

Table 17: Descriptive statistics of the data set generated during leave-one-out CV

Parameter	Unit	N	Mean	Median	Standard Deviation	Coefficient of Variation
ψ_a	cm	115	17.5	17.5	0.621	3.55e-2
θ_s	cm^3/cm^3	115	0.374	0.372	2.77e-2	7.39e-2

Table 18: SW *W* statistic and significance levels for the CV data set

Parameter	Unit	N	W
ψ_a	cm	115	0.989
θ_s	cm^3/cm^3	115	0.964 **
$1/\theta_s$	cm^3/cm^3	115	0.982

Table 19: Trend surface analysis results conducted for the CV data set to determine if the parameter contains a geographic trend

Parameter	Unit	N	R^2	F-statistic
ψ_a	cm	115	0.603	33.1 ***
θ_s	cm^3/cm^3	115	0.562	35.3 ***

Table 20: t-values used to identify specific trend directions for the CV data set

Parameter	Unit	N	x	y	x*y	x^2	y^2
ψ_a	cm	115	-7.20 ***	NS	7.77 ***	4.27 ***	-3.06 **
θ_s	cm^3/cm^3	115	-11.6 ***	NS	NS	11.2 ***	NS

Table 21: Descriptive statistics of trend surface residuals computed for the CV data set

Parameter	Unit	N	Mean	Median	Standard Deviation
ψ_a	cm	115	2.09e-18	2.78e-2	0.391
θ_s	cm^3/cm^3	115	4.44e-18	2.89e-3	1.91e-2

Table 22: SW *W* statistic and significance level for the CV data set residuals

Parameter	Unit	N	W
ψ_a	cm	115	0.984
θ_s	cm^3/cm^3	115	0.976 *

Table 23: Cross-variogram with spherical model computed for the residual values of ψ_a and θ_s for the CV data set

Parameters	Unit	Model	Nugget	Sill	Range (mm)	R^2	NRMSE	DSD
ψ_a vs. θ_s	-	Spherical	-2.34e-3	1.42e-2	15.5	0.897	0.109	0.165

Table 24: CV results using CK with the spherical model for ψ_a of the CV data set

Kriging Method	Model	r	NRMSE	a	b
CK	Spherical	0.824	0.102	7.076	0.595

Table 25: CV results using CK with the spherical model for θ_s of the CV data set

Kriging Method	Model	r	NRMSE	a	b
CK	Spherical	0.716	0.126	0.212	0.434

Table 26: Descriptive statistics of the detrended variable, z_{ra} (i.e. presented as "Krige" in tables 27 and 28 below

Parameter	Unit	N	Mean	Median	Standard Deviation	Coefficient of Variation
ψ_a	cm	115	17.4	17.4	0.564	3.23e-2
θ_s	cm^3/cm^3	115	0.373	0.366	3.65e-2	9.77e-2

Table 27: MW *W* statistic with significance levels to determine statistical differences between the following data sets: Cropper (2014), Kang et al. (2014), corrected z_{ra} (labeled as Krige) variable, and Trend (created during trend surface analysis)

Cropper vs.	ψ_a	θ_s	λ	θ_r
Kang	227 **	327 *	384	1.08 ***
Krige	197 ***	290 **	NA	NA
Trend	206 ***	307 *	206	1.08 ***

Table 28: KS *D* statistic with significance levels to determine statistical differences between the following data sets: Cropper (2014), Kang et al. (2014), corrected z_{ra} (labeled as Krige) variable, and Trend (created during trend surface analysis)

Cropper vs.	ψ_a	θ_s	λ	θ_r
Kang	0.673 **	0.595 **	0.434	0.656 **
Krige	0.717 ***	0.613 **	NA	NA
Trend	0.708 ***	0.604 **	0.469	0.691 ***

Appendix 2. Figures

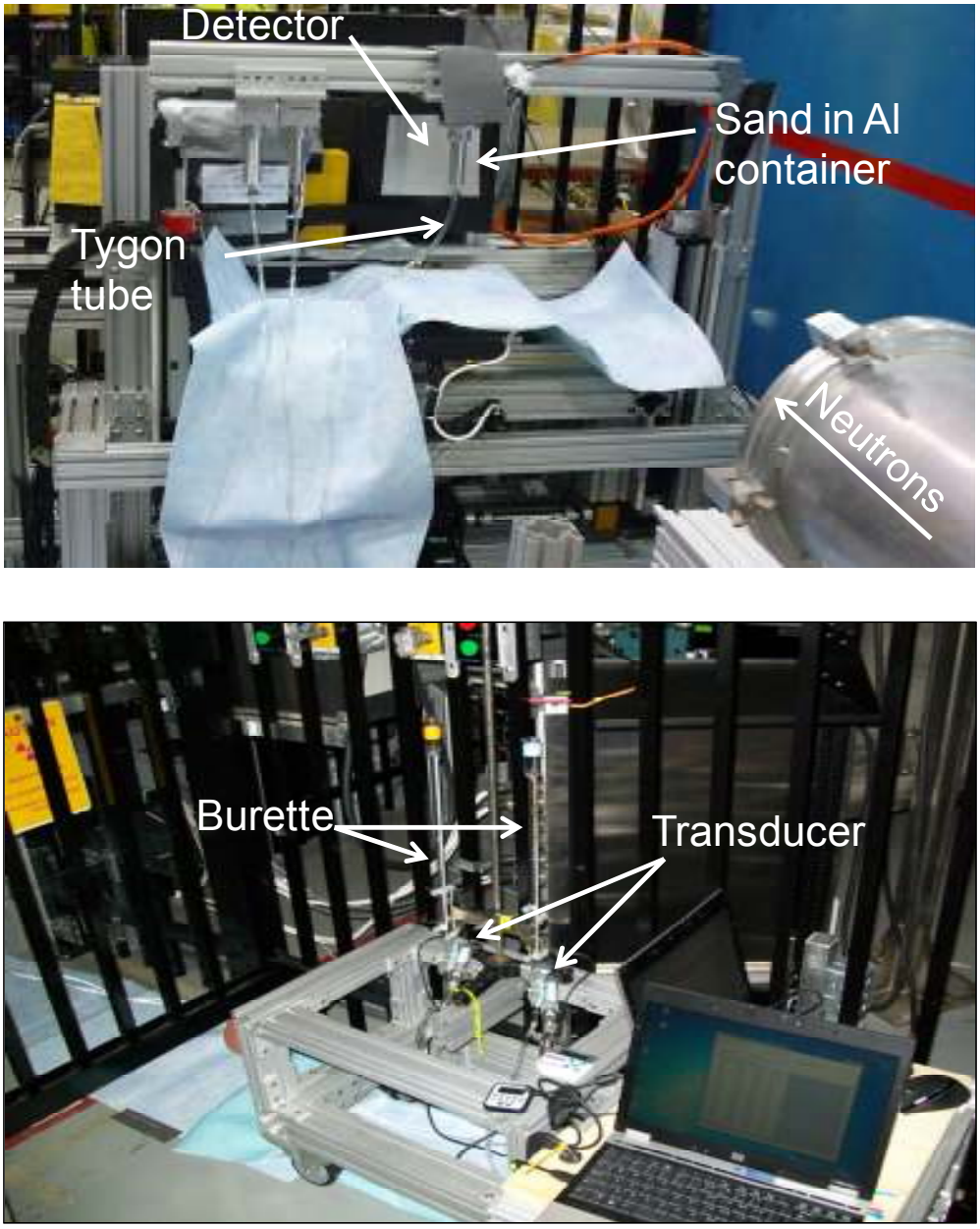


Figure 1: a) Hanging water column set up for drainage of Flint sand at the HFIR CG1-D beamline. b) Various basal matric potentials were achieved by adjusting the height of the burette column outside of the beamline (Kang et al., 2014).

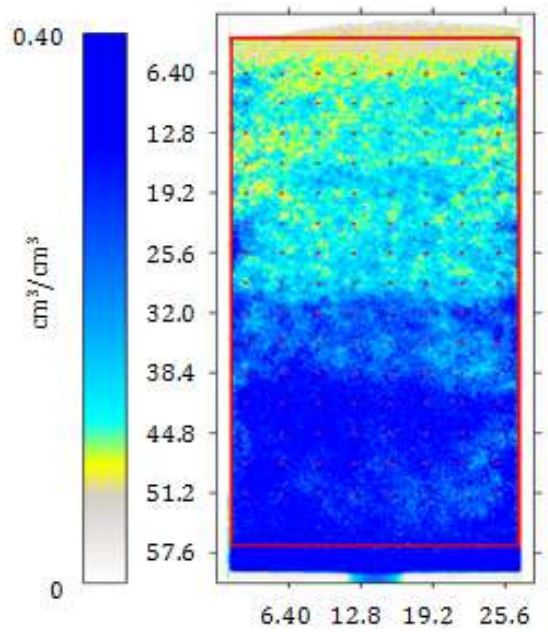


Figure 2: The 8 x 15 grid superimposed on an example 2-D radiographic image (28.2 x 60.7 mm) of air (white) displacing water (blue) in Flint sand at a given equilibrium basal matric potential used to construct the 120 point water retention curves (Kang et al., 2014); the distance between adjacent points is 3.2 mm, and the dimensions of the red rectangular region of interest are 25.6 x 51.2 mm

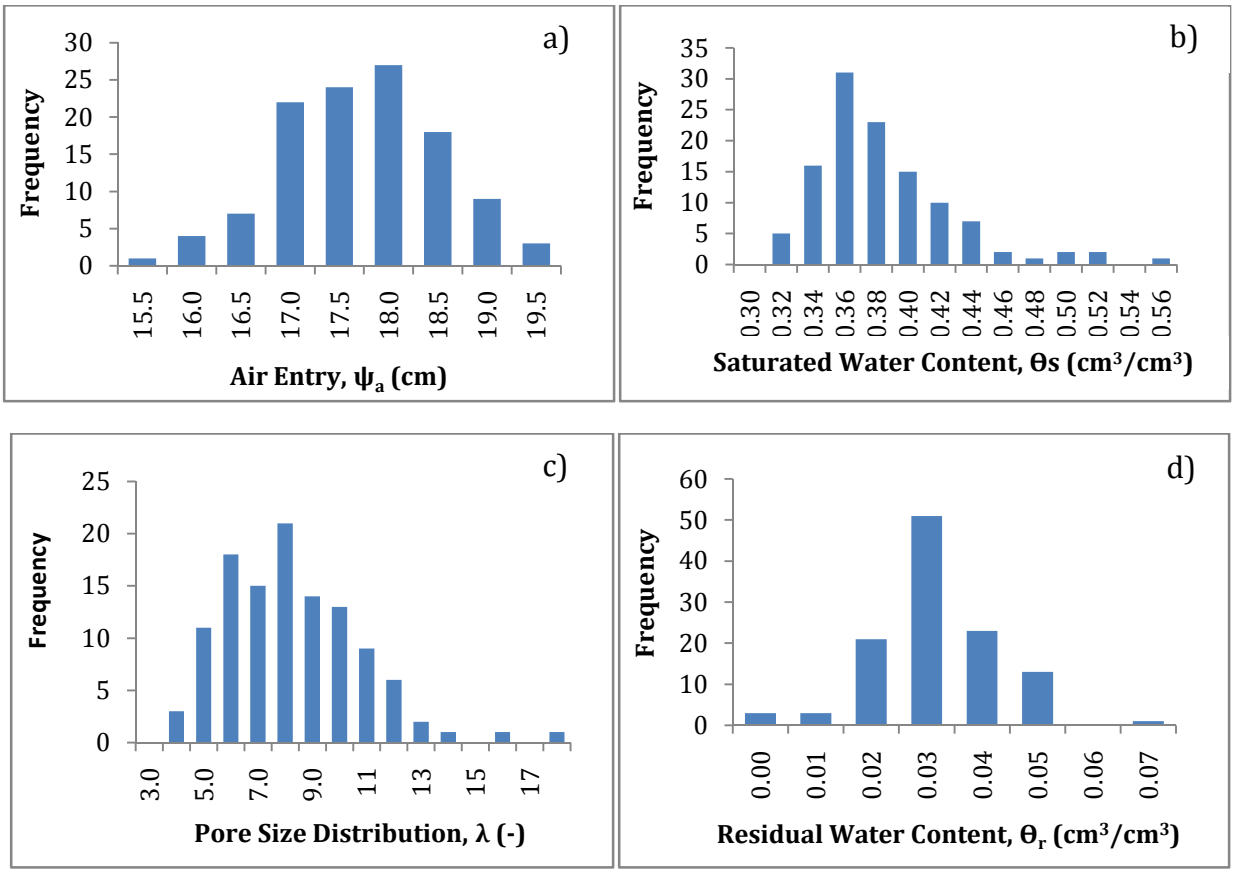
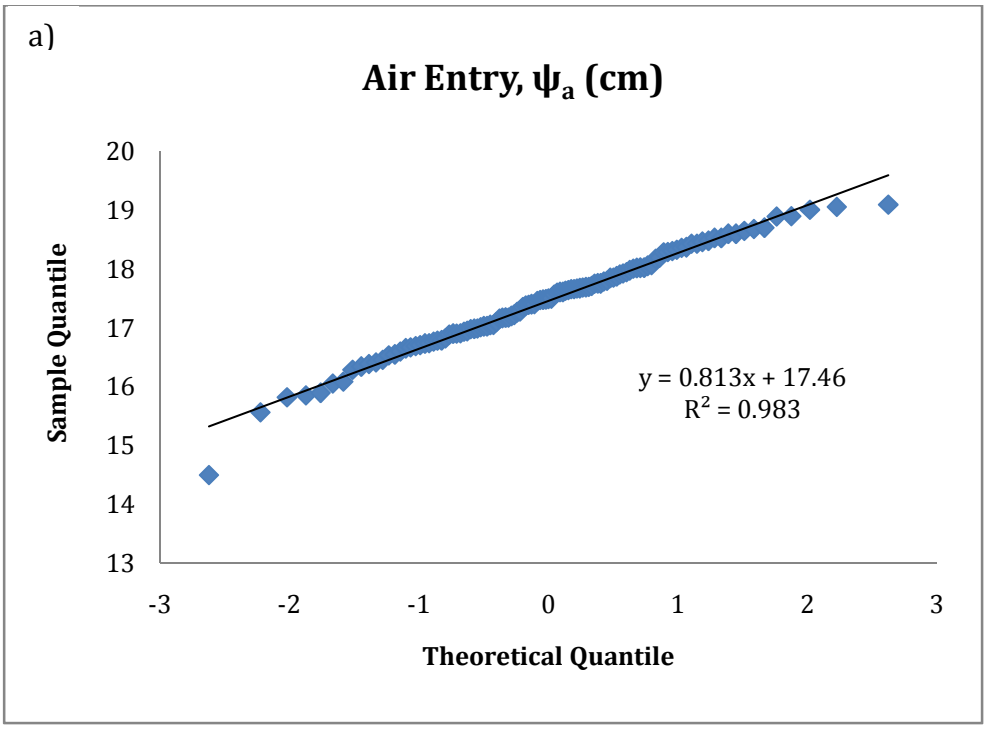


Figure 3: Histograms of the observed BC parameters used to indicate that sub-core scale heterogeneity is present

Figure 4: Q-Q plots of the BC parameters used to visualize the SW results given in Table 3



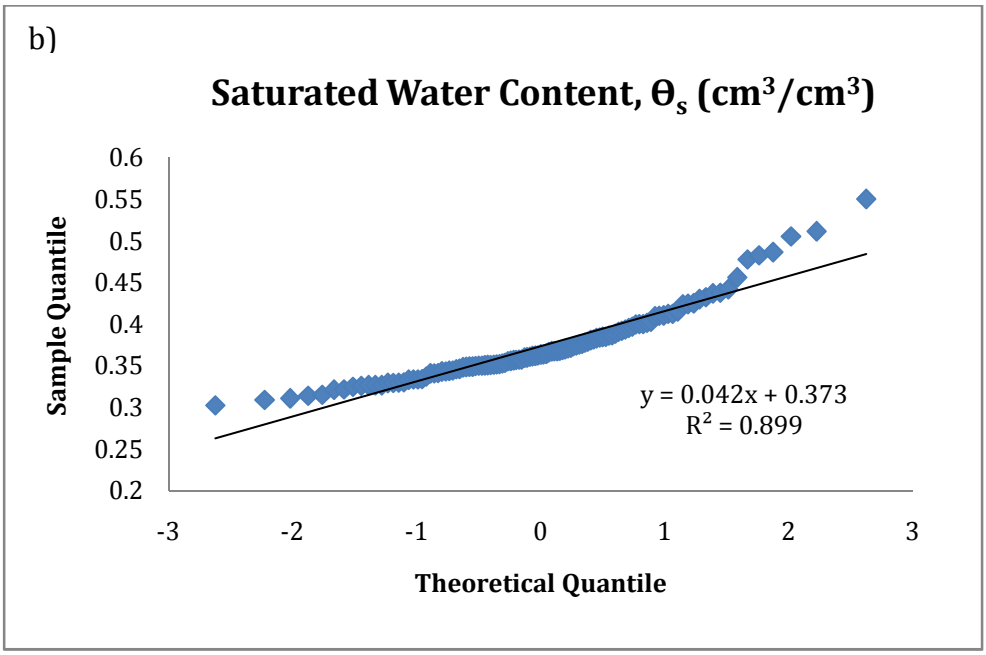
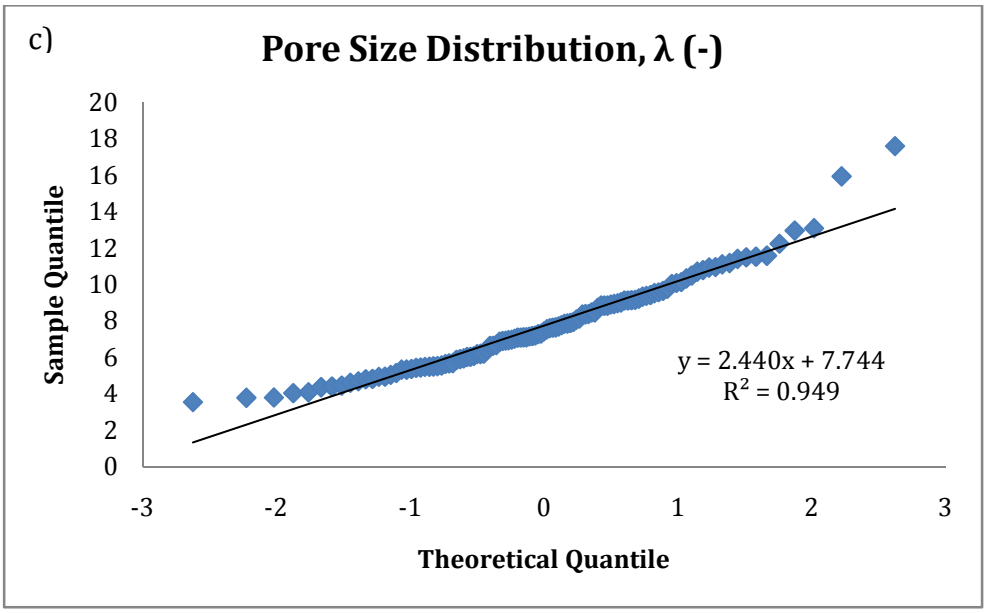


Figure 4 continued



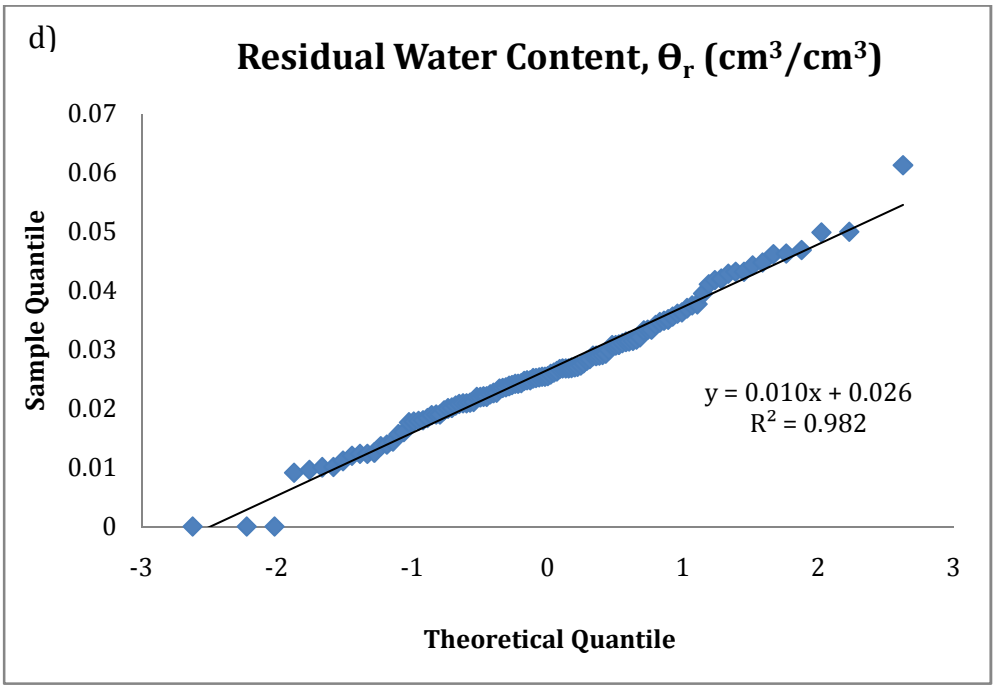
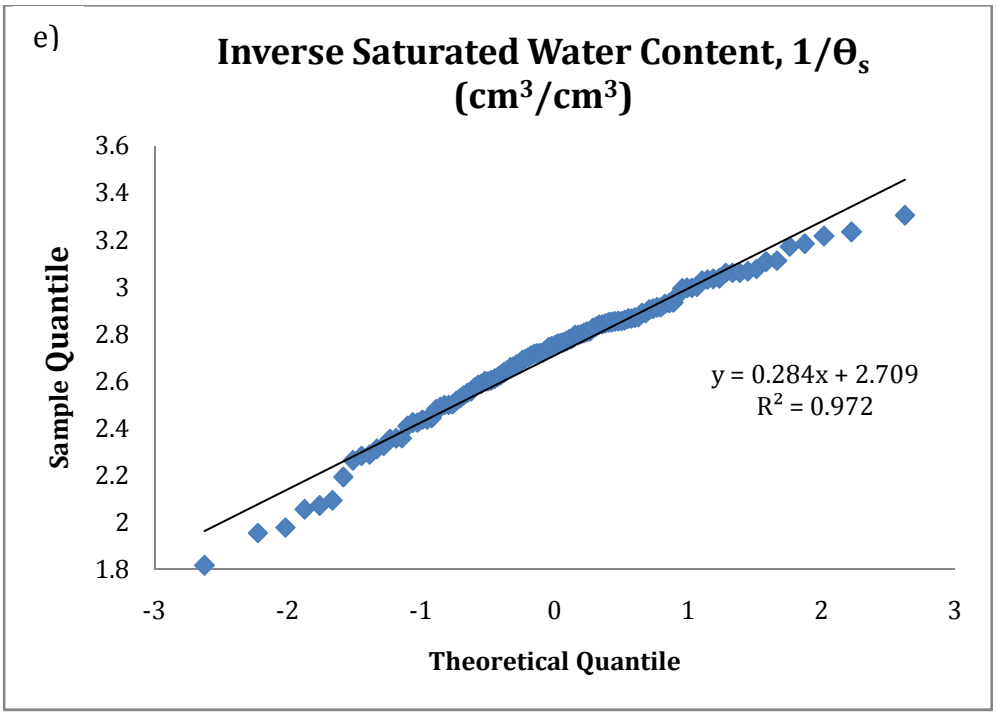


Figure 4 continued



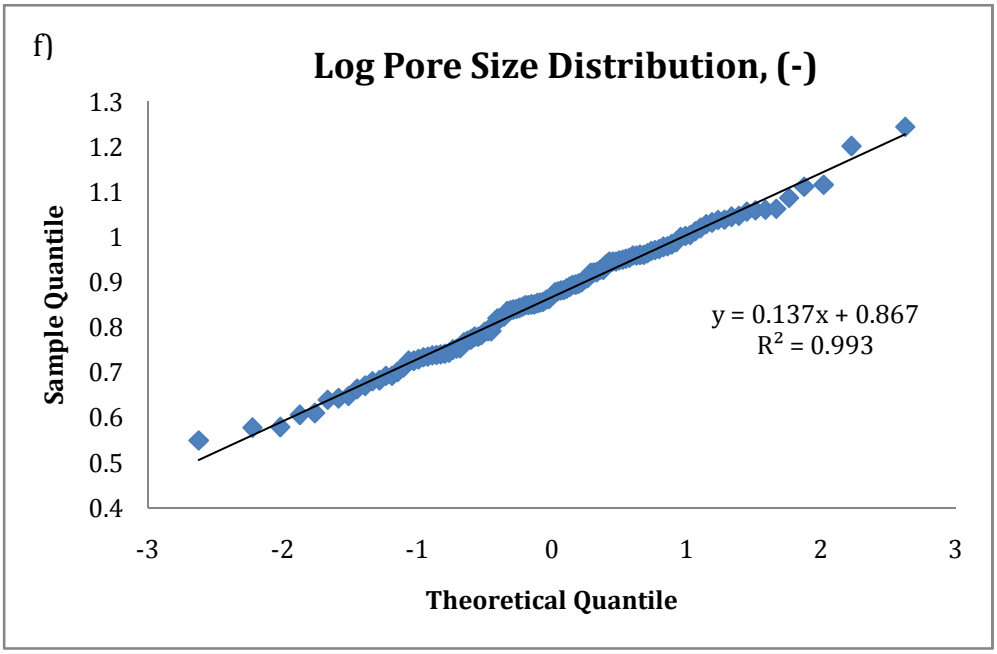
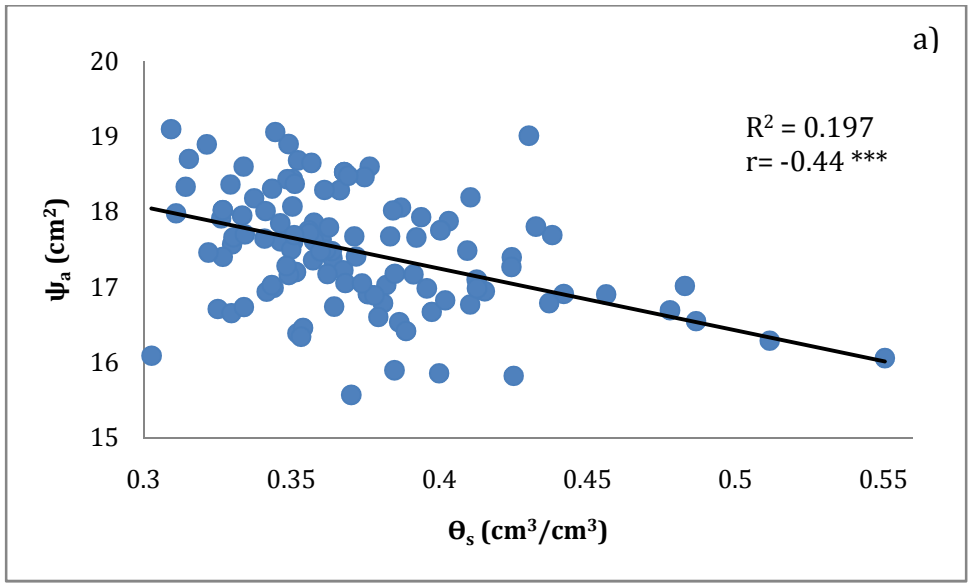


Figure 4 continued

Figure 5: Scatter diagrams used to indicate significant linear relationships between the BC parameters.



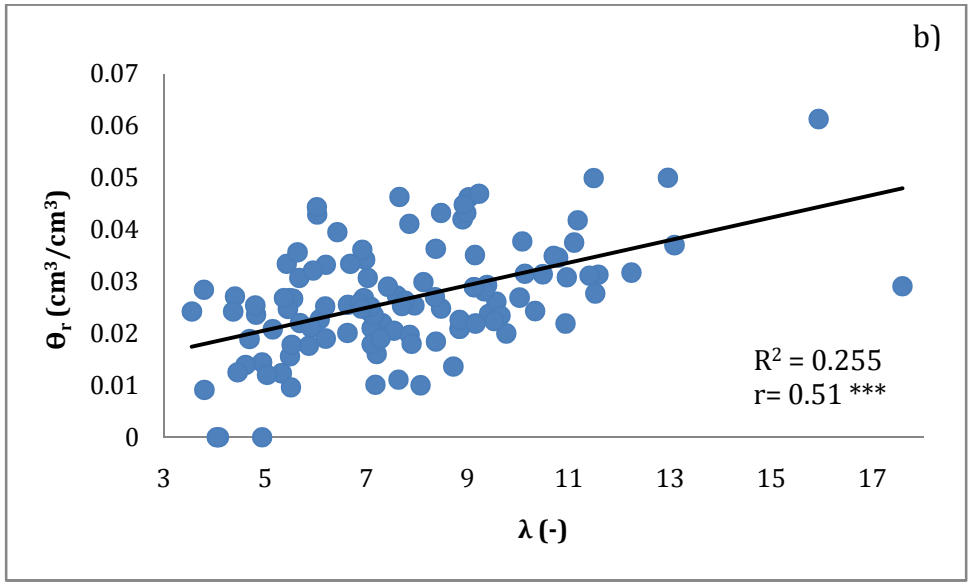
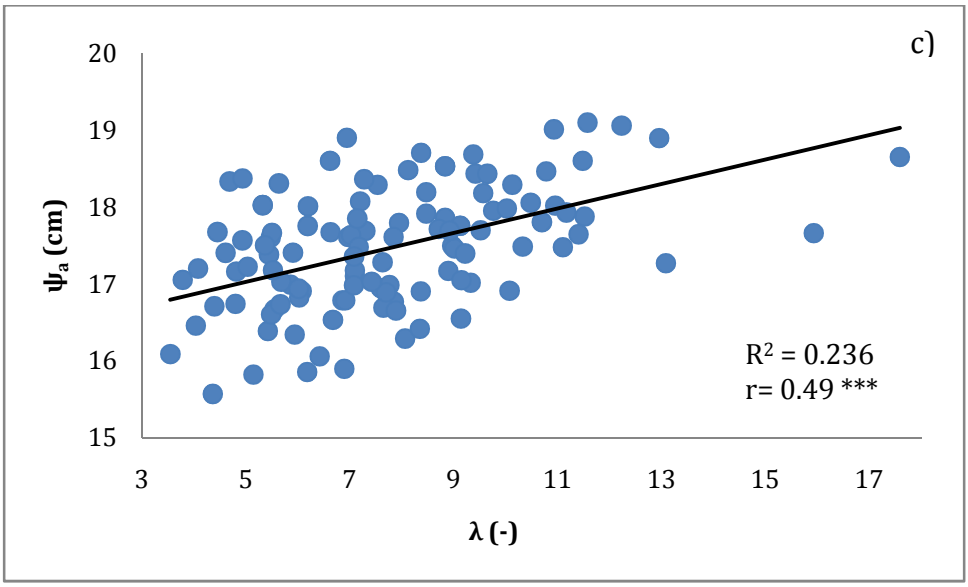


Figure 5 continued



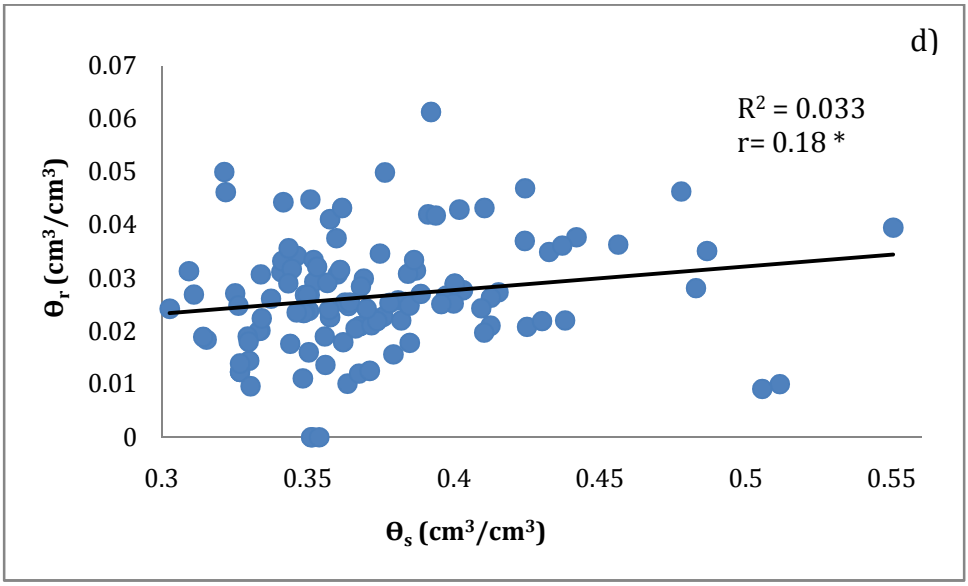


Figure 5 continued

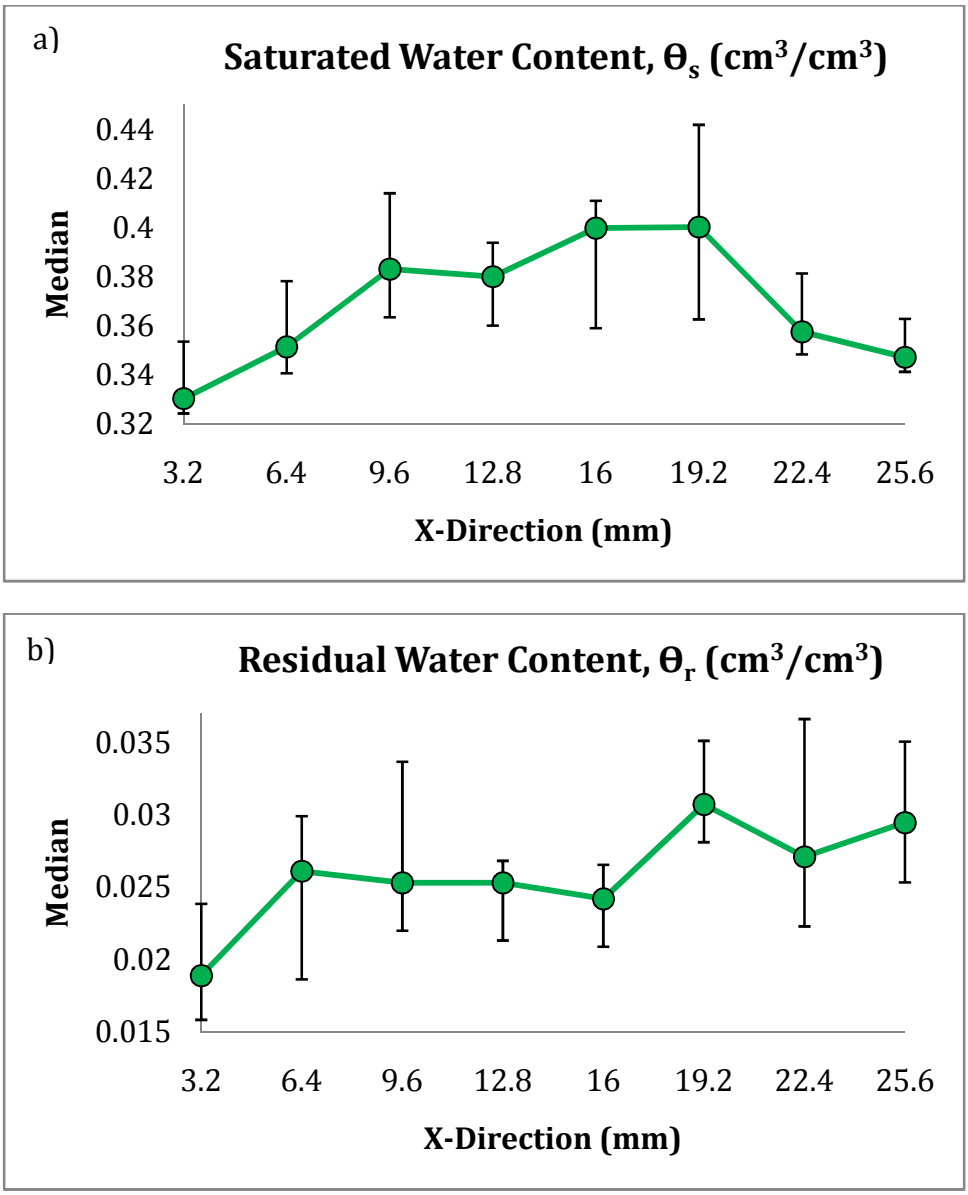


Figure 6: Plots of median values and their corresponding upper (75%) and lower (25%) quartiles for each sample interval (i.e. 3.2 mm, 6.4 mm, etc) in the x-direction. Since water retention curves were computed for all but five points on an 8 x 15 Cartesian grid, sample intervals in the x-direction typically consist of 8 values. These plots are used to illustrate KW results and determine significantly different sample intervals. BC parameters with significantly different median values in the x-direction are shown here, for θ_s and θ_r.

Figure 7: Plots of median values and their corresponding upper (75%) and lower (25%) quartiles for each sample interval (i.e. 6.4 mm, 9.6 mm, etc) in the y-direction. Since water retention curves were computed for all but five points on an 8 x 15 Cartesian grid, sample intervals in the y-direction typically consist of 15 values. These plots are used to illustrate the KW results and determine significantly different sample intervals. BC parameters with significantly different median values in the y-direction are shown here, for ψ_a and λ.

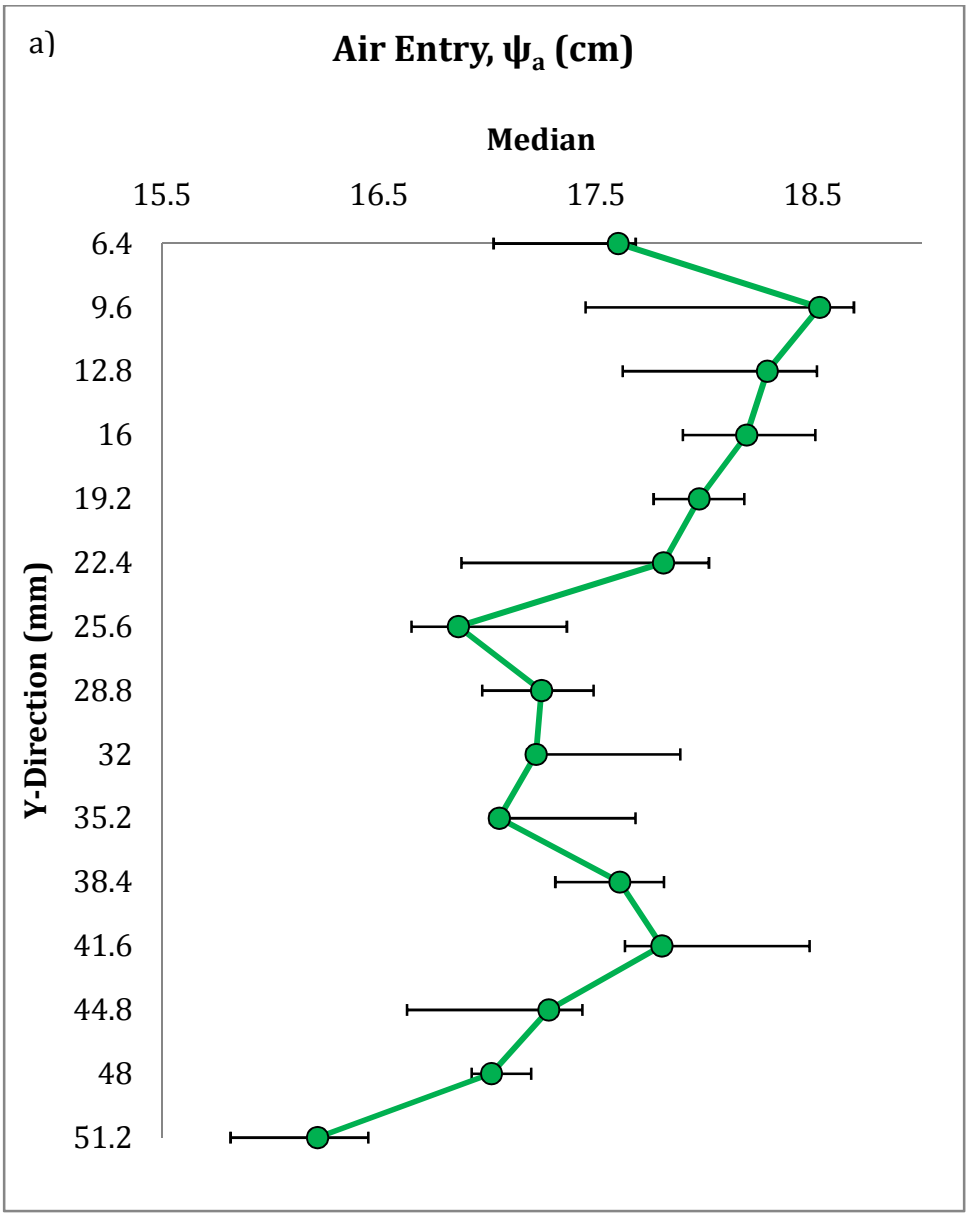


Figure 7 continued

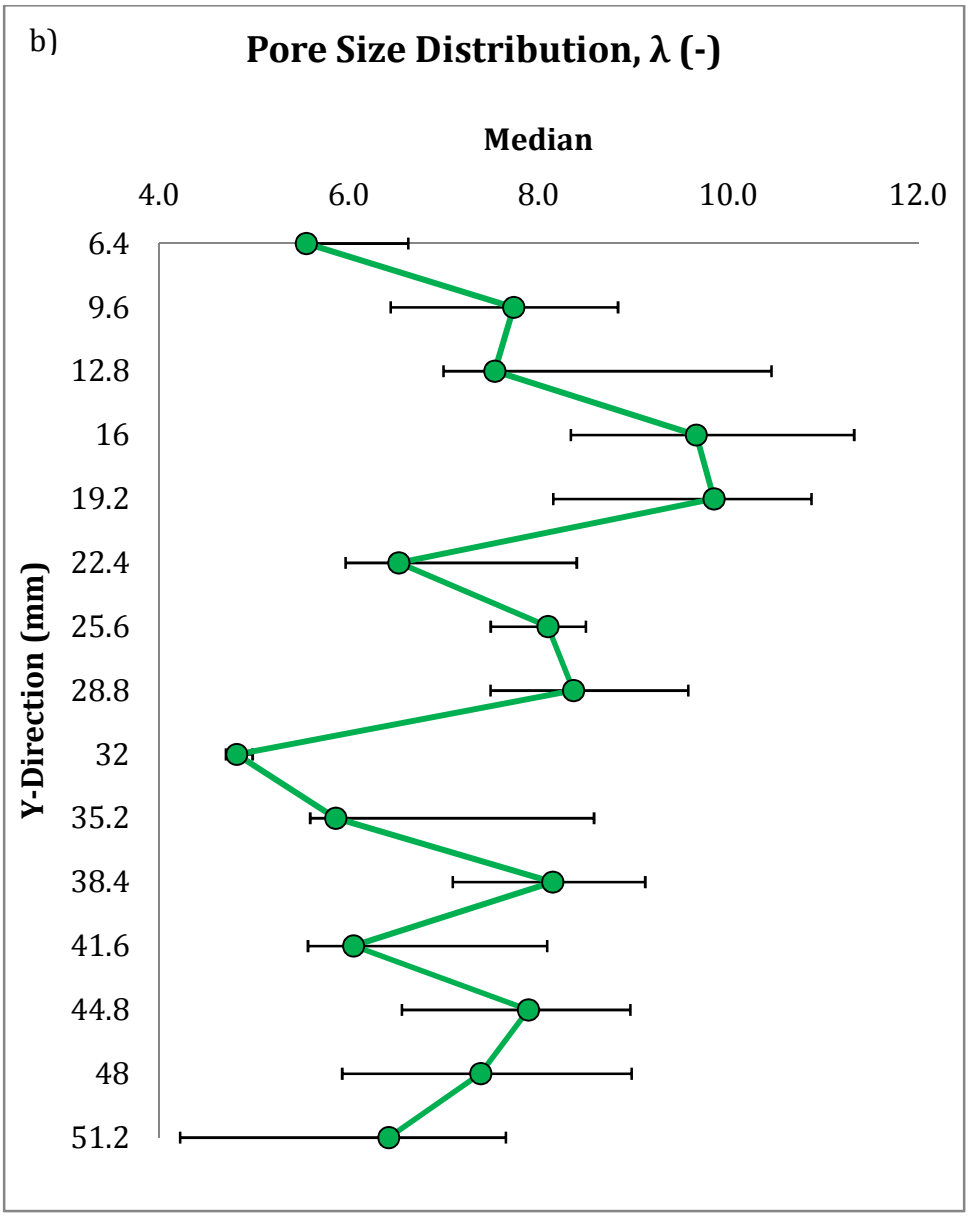


Figure 7 continued

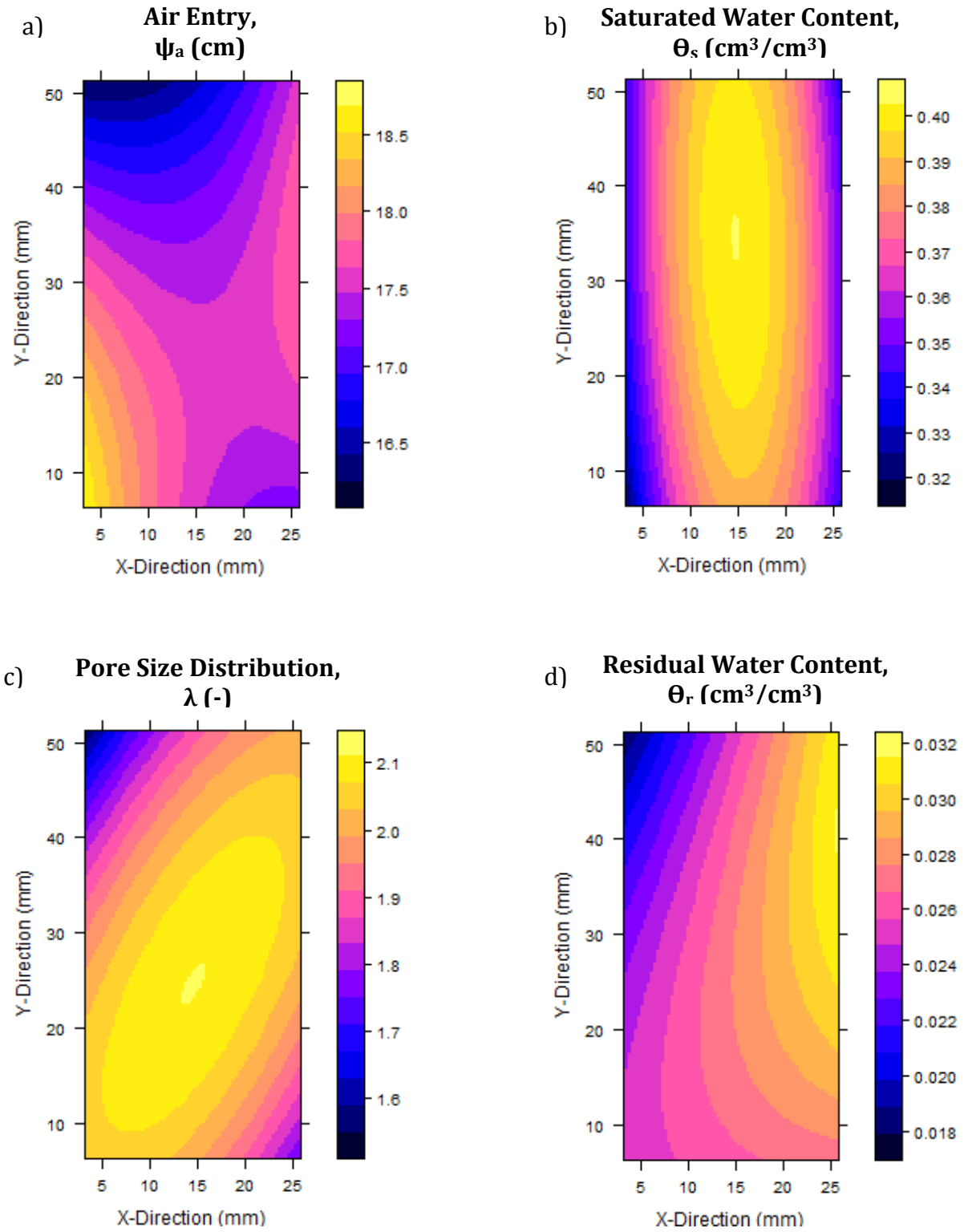
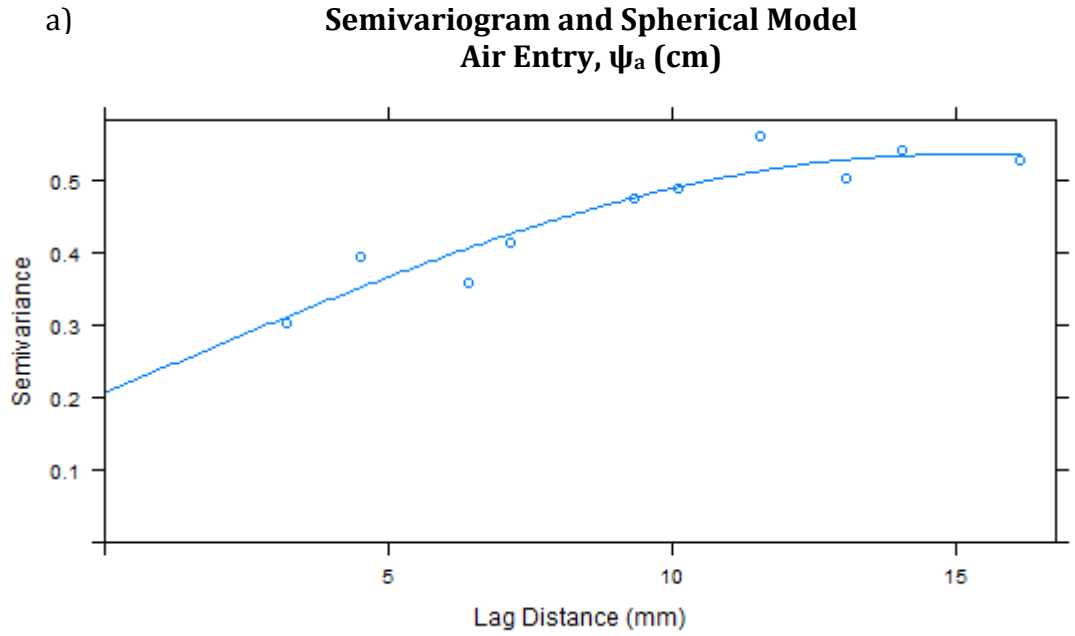


Figure 8: Trend surfaces for the four BC parameters are shown here and were used to identify 2 dimensional trends

Figure 9: Residual semivariograms and best fit spherical model for all four BC parameters, a) ψ_a and b) θ_s were found to be spatially dependent whereas c) λ and d) θ_r were found to be pure nugget effect

a)



b)

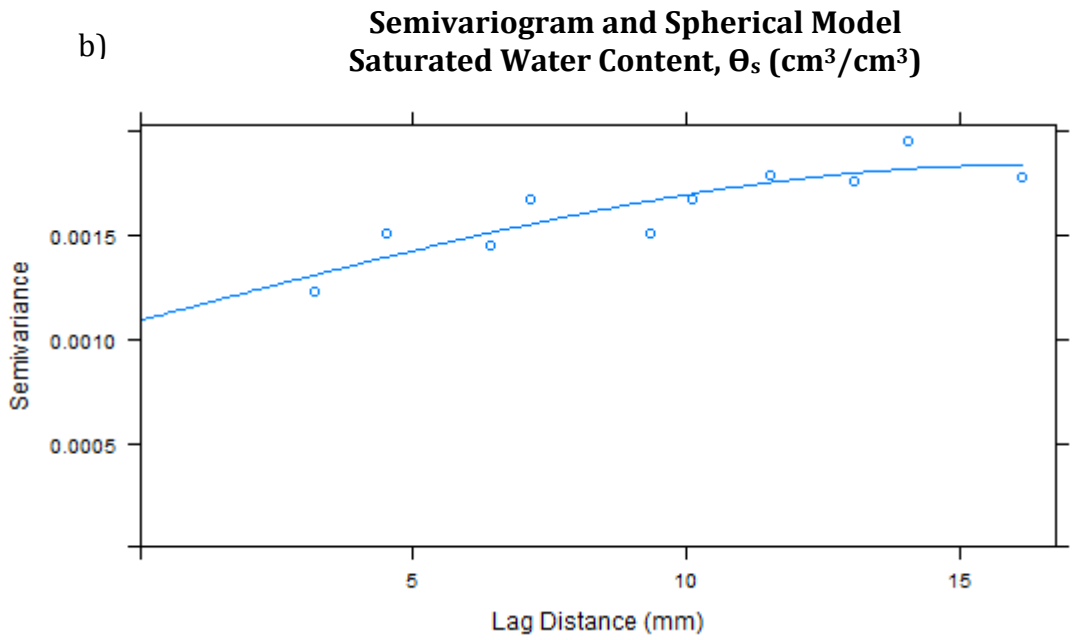
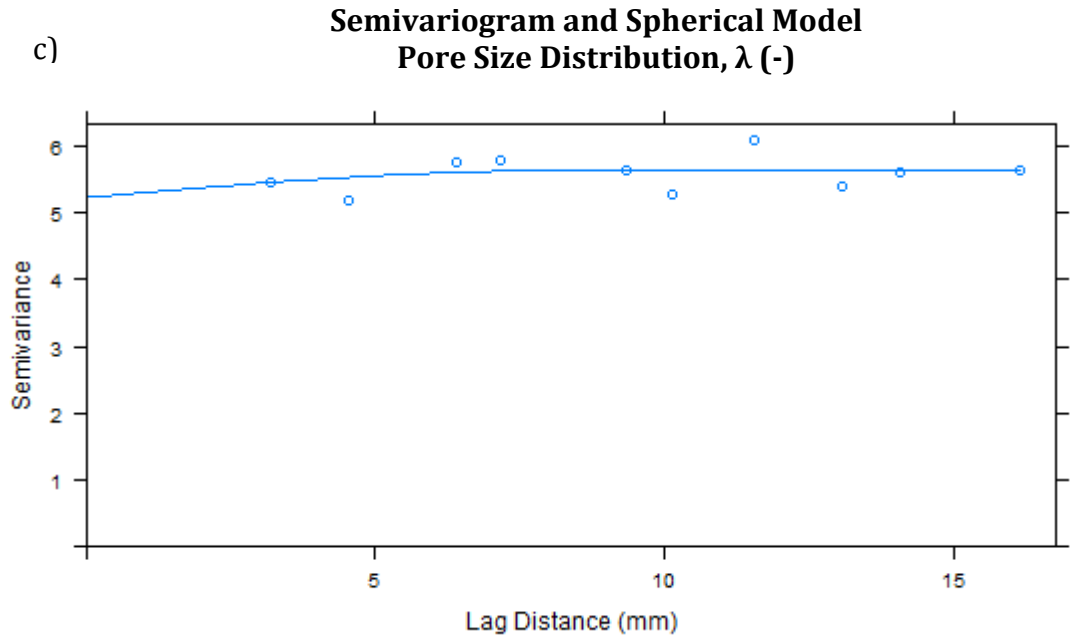


Figure 9 continued

c)



d)

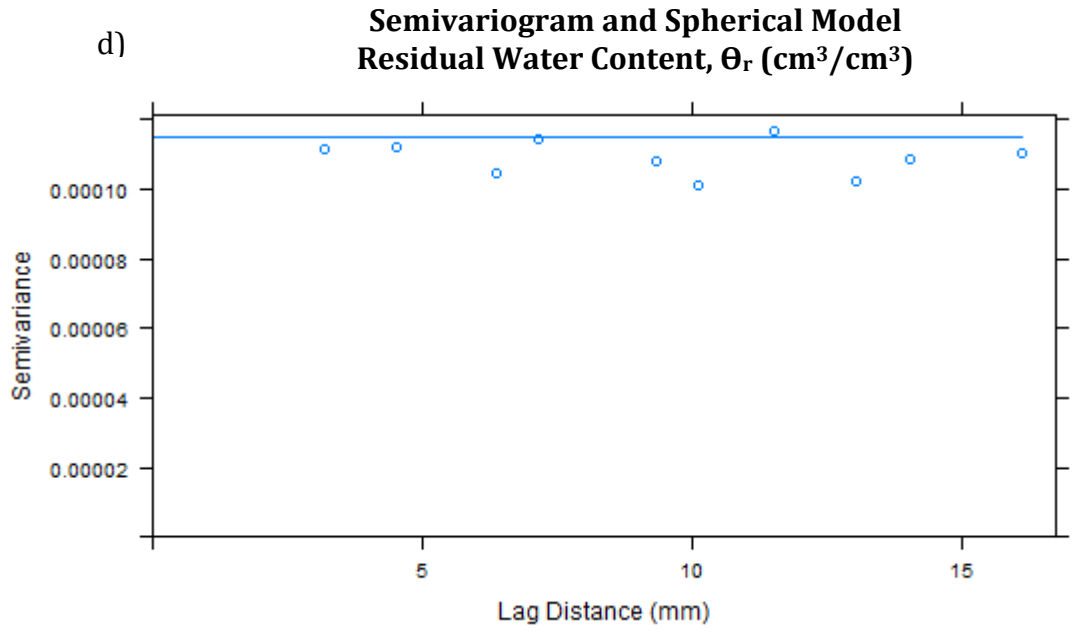
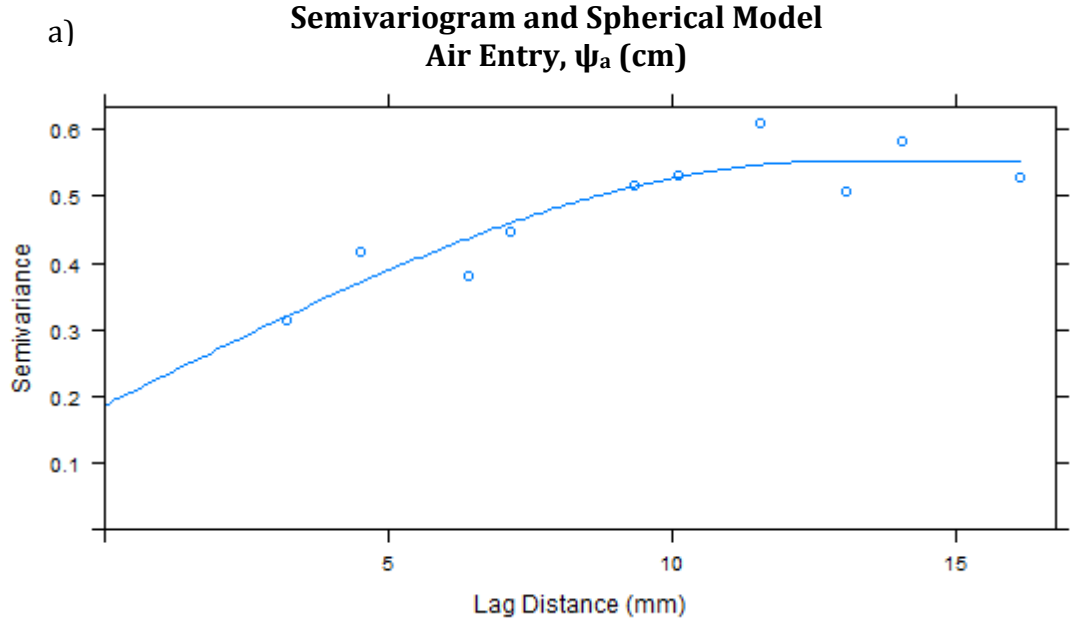


Figure 9 continued

Figure 10: Raw semivariograms and best fit spherical model for ψ_a, θ_s, and λ parameters, a) ψ_a and b) θ_s were found to be spatially dependent whereas c) λ was determined to be pure nugget effect; a significant fit of the spherical model to the semivariogram of θ_r was unsuccessful

a)



b)

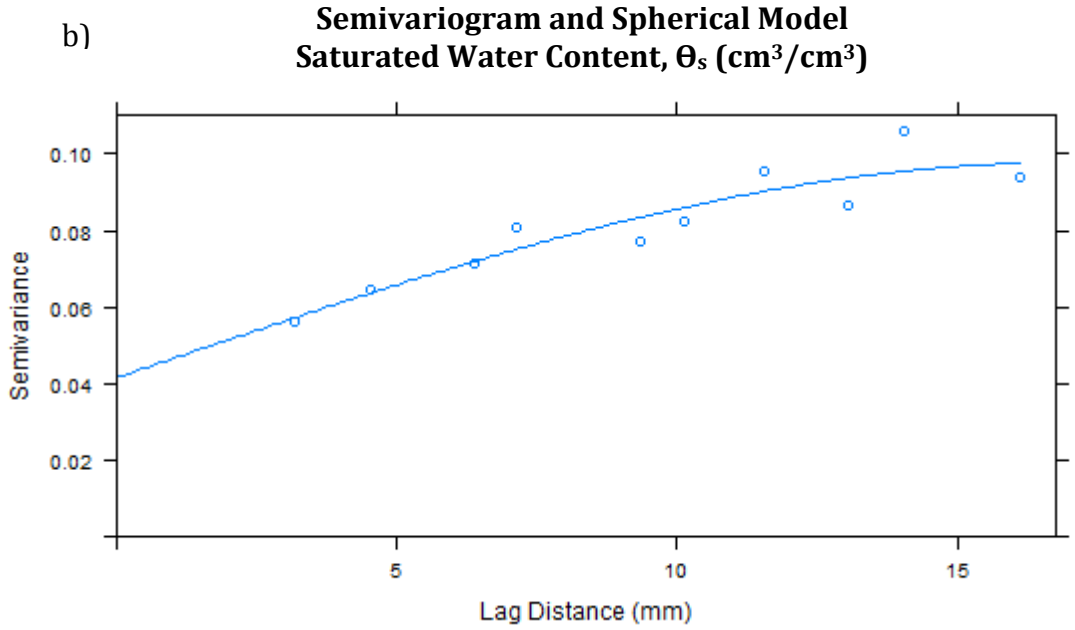


Figure 10 continued

c)

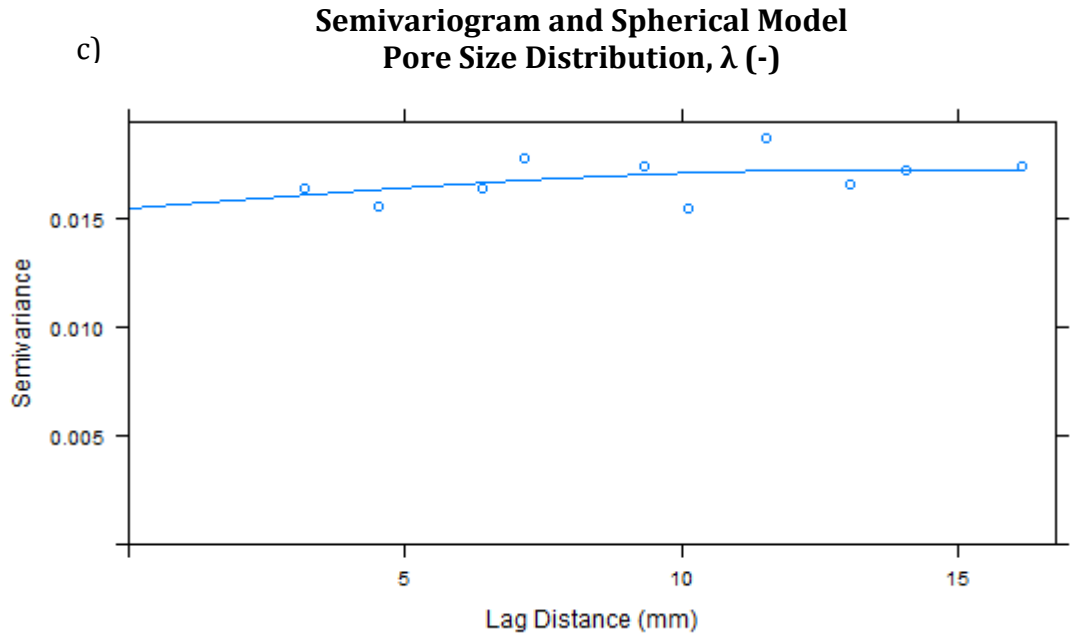


Figure 10 continued

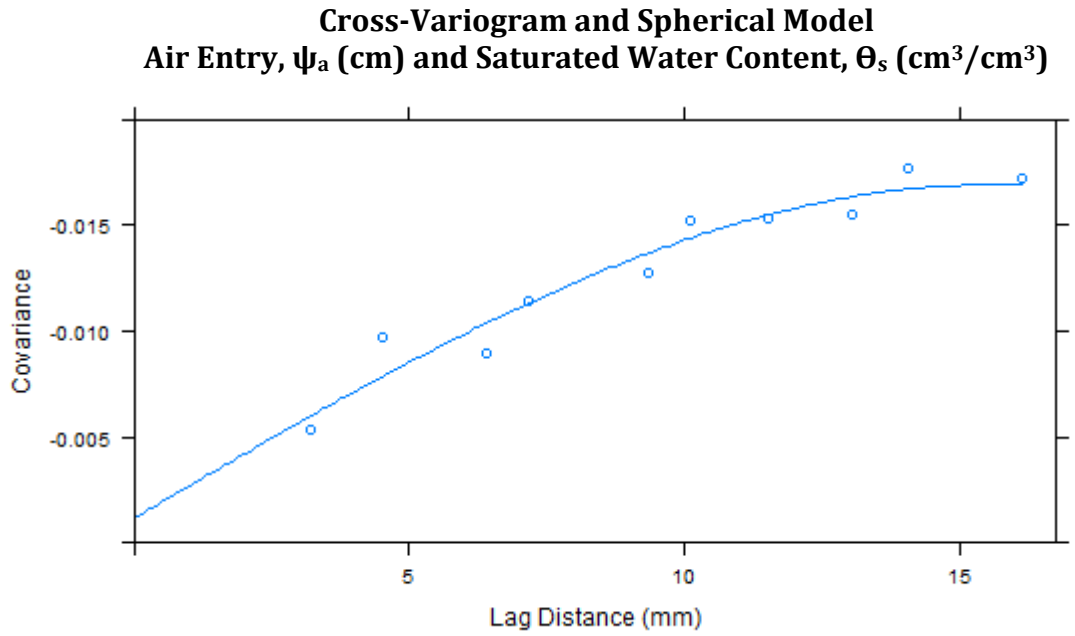


Figure 11: Shows the cross-variogram of ψ_a and θ_s with the fitted spherical model

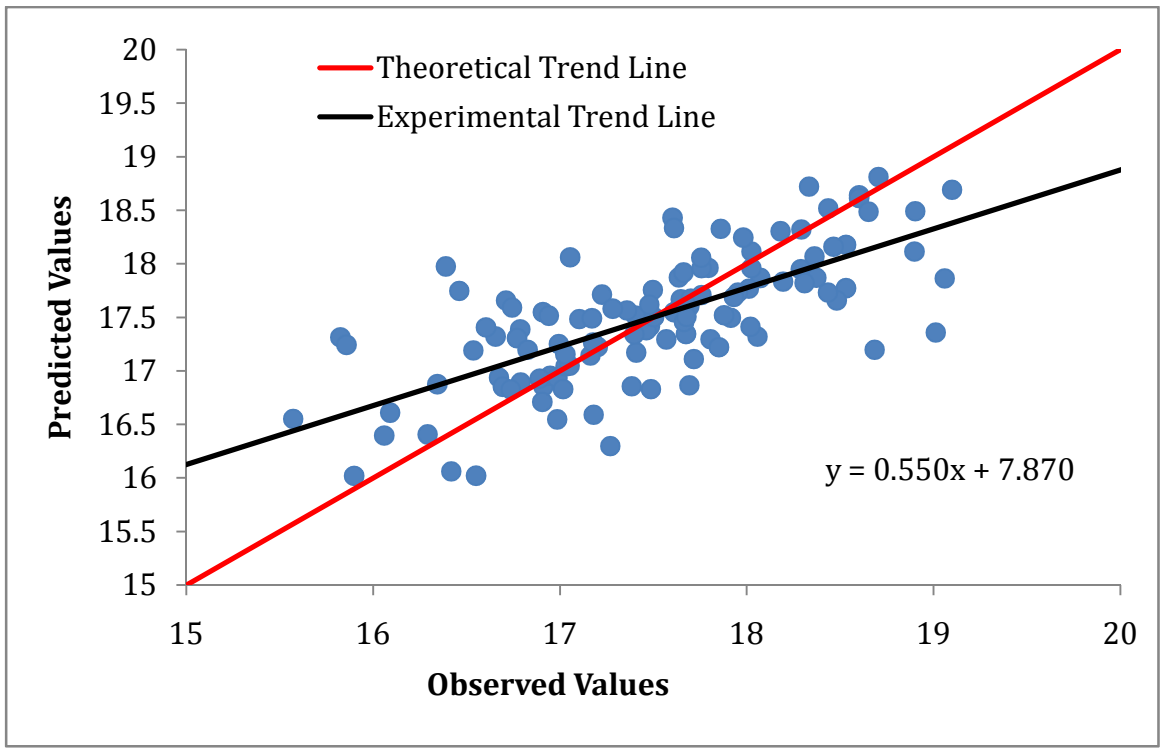


Figure 12: One-to-one graph of the CK method compared to the theoretical one-to-one line of correlation for ψ_a

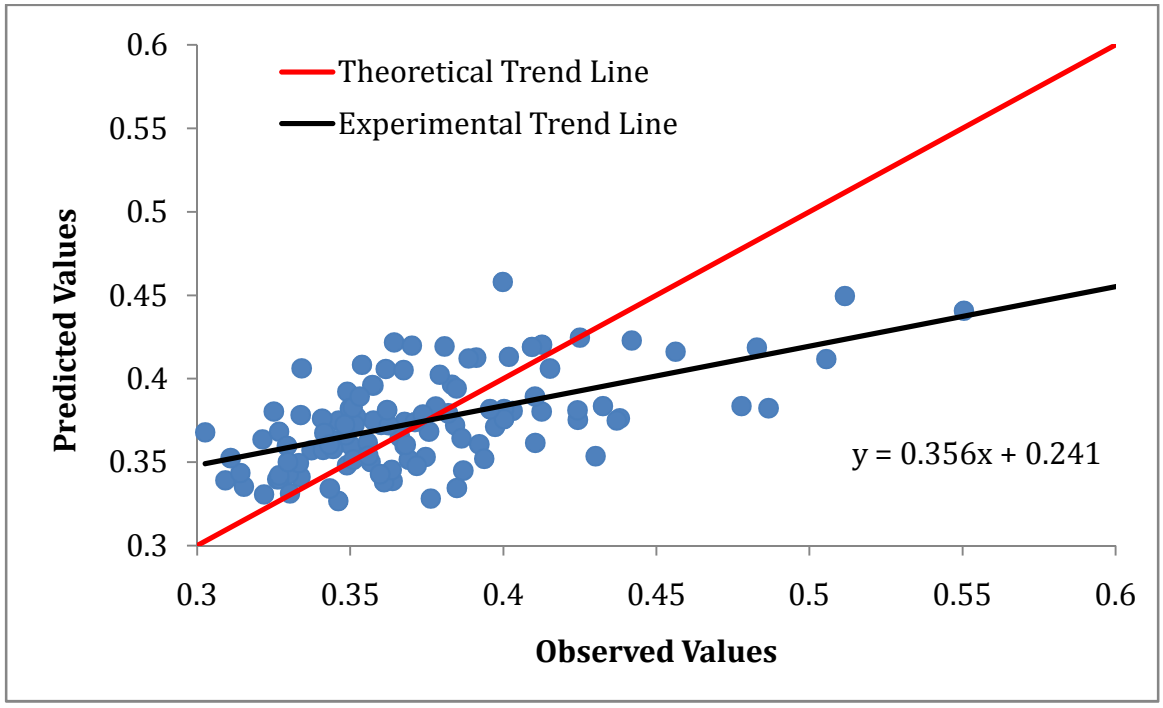


Figure 13: One-to-one graph of the CK method compared to the theoretical one-to-one line of correlation for θ_s

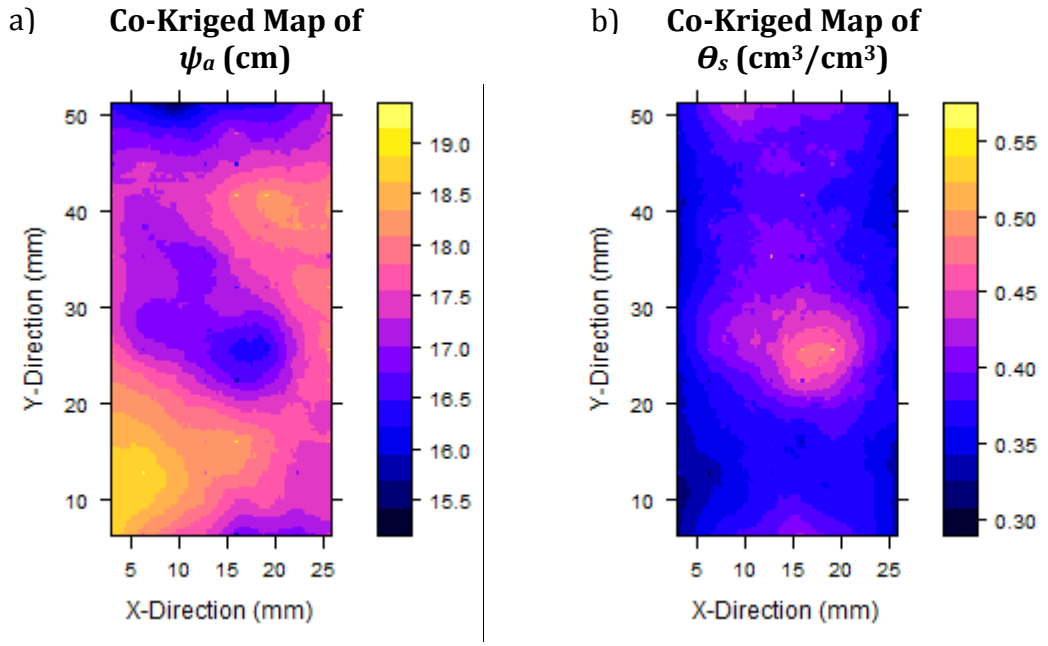


Figure 14: Final co-kriged maps for ψ_a and θ_s using the Kang et al. (2014) data set

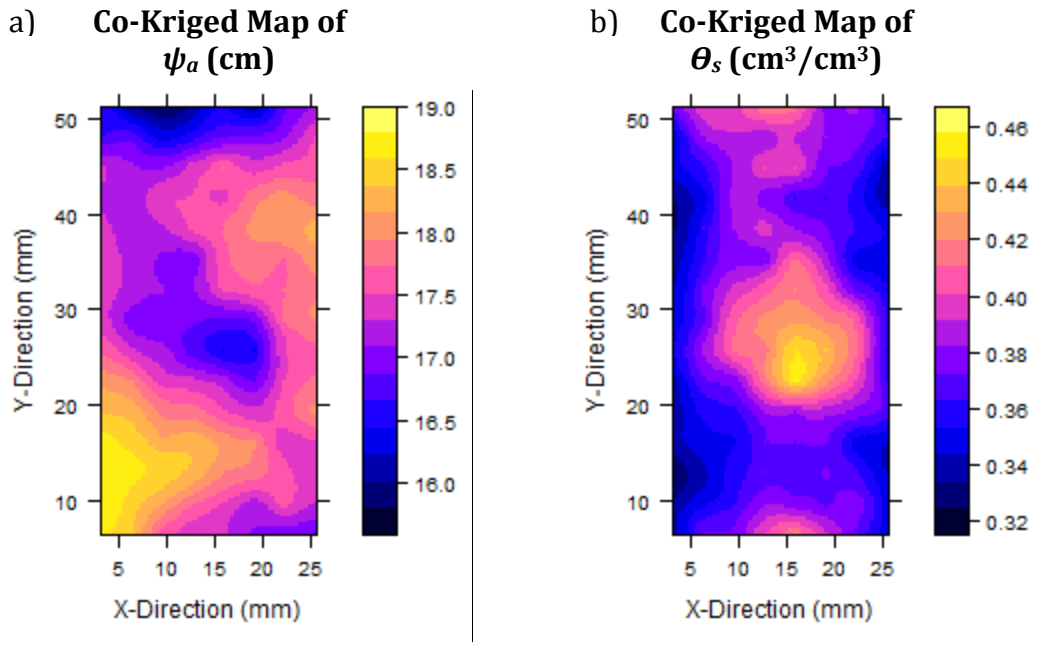


Figure 15: Final co-kriged maps for ψ_a and θ_s using the CV data set

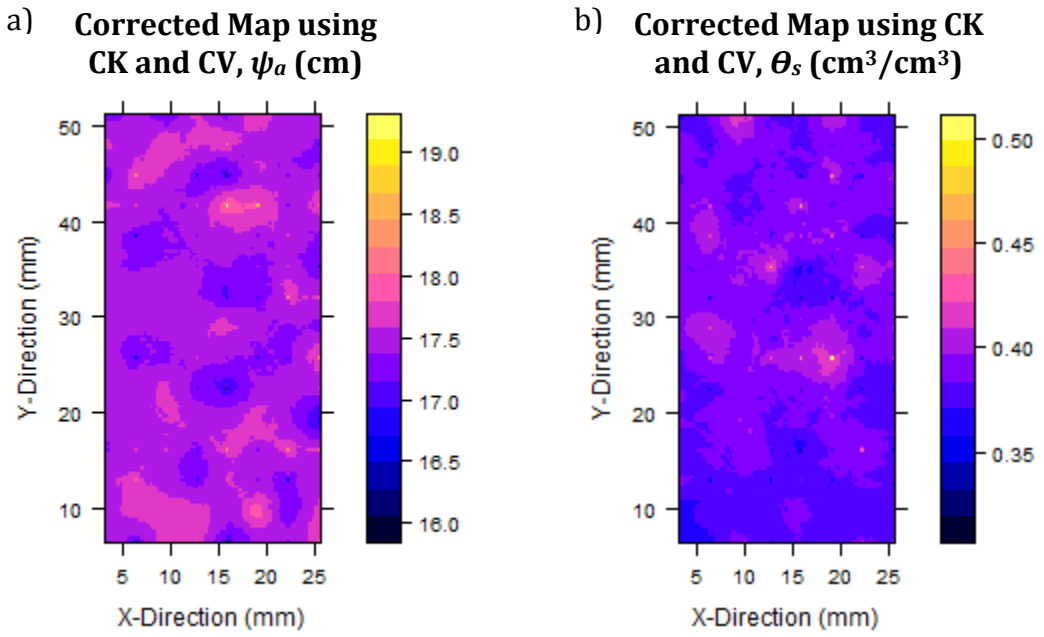
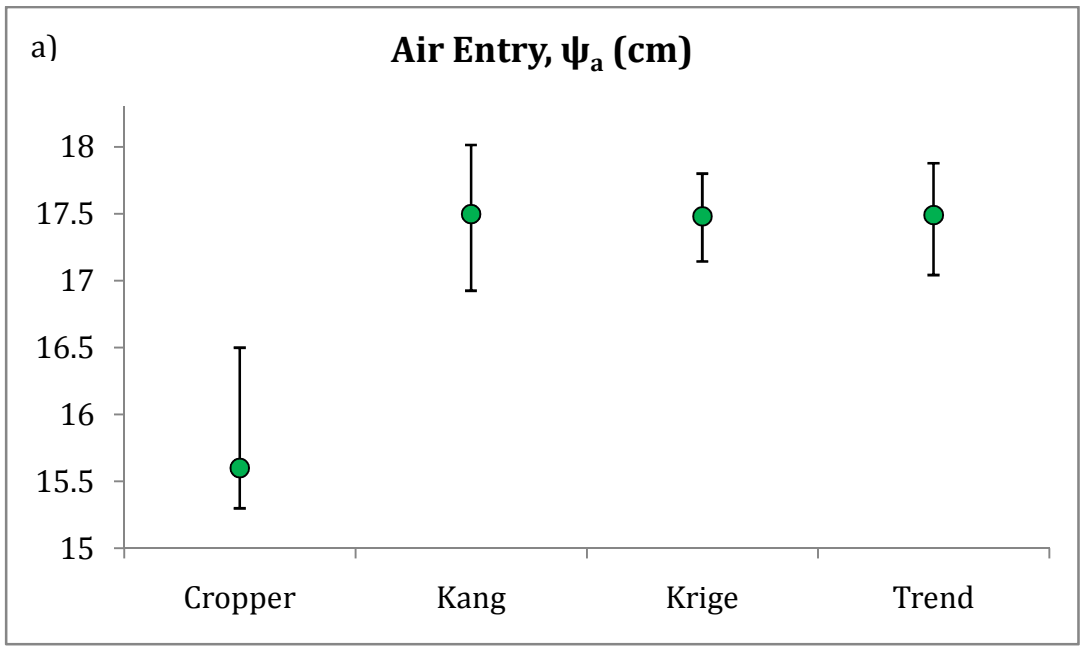


Figure 16: Corrected maps for ψ_a and θ_s using the cross-validation and co-kriging correction method

Figure 17: Plots of median values and their corresponding upper (75%) and lower (25%) quartiles for the independent Cropper (2012) data set, the Kang et al. (2014) data set, the Krige data set (i.e. spatial trend corrected data using CK and CV for ψ_a and θ_s), and the Trend data set (i.e. trend corrected data using trend surface analysis for all BC parameters). These plots are used to illustrate MW and KS results and determine which data sets are significantly different from one another.



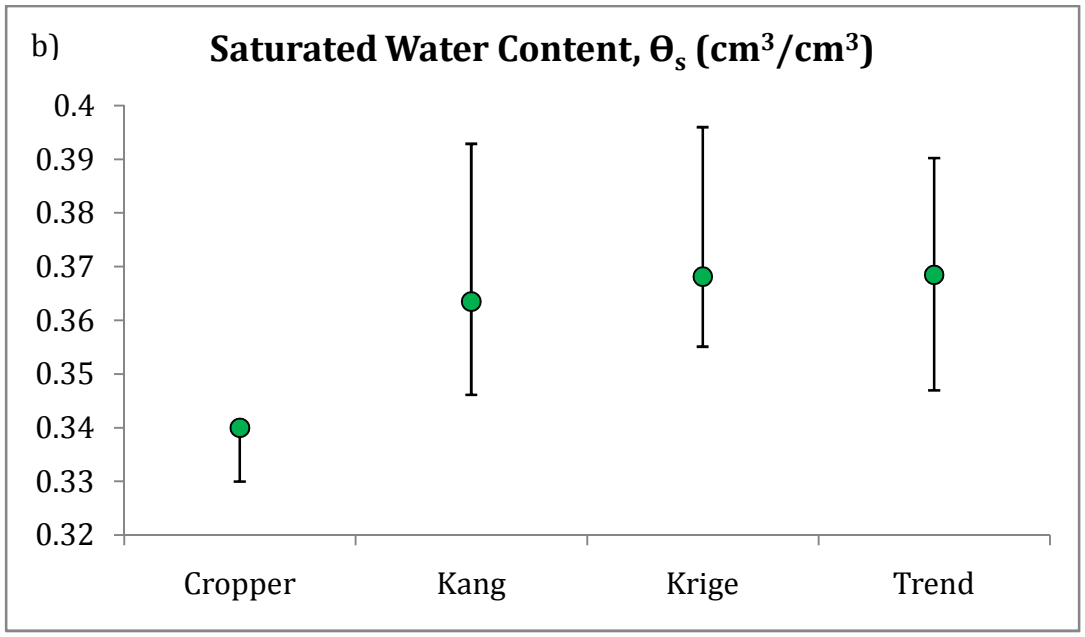
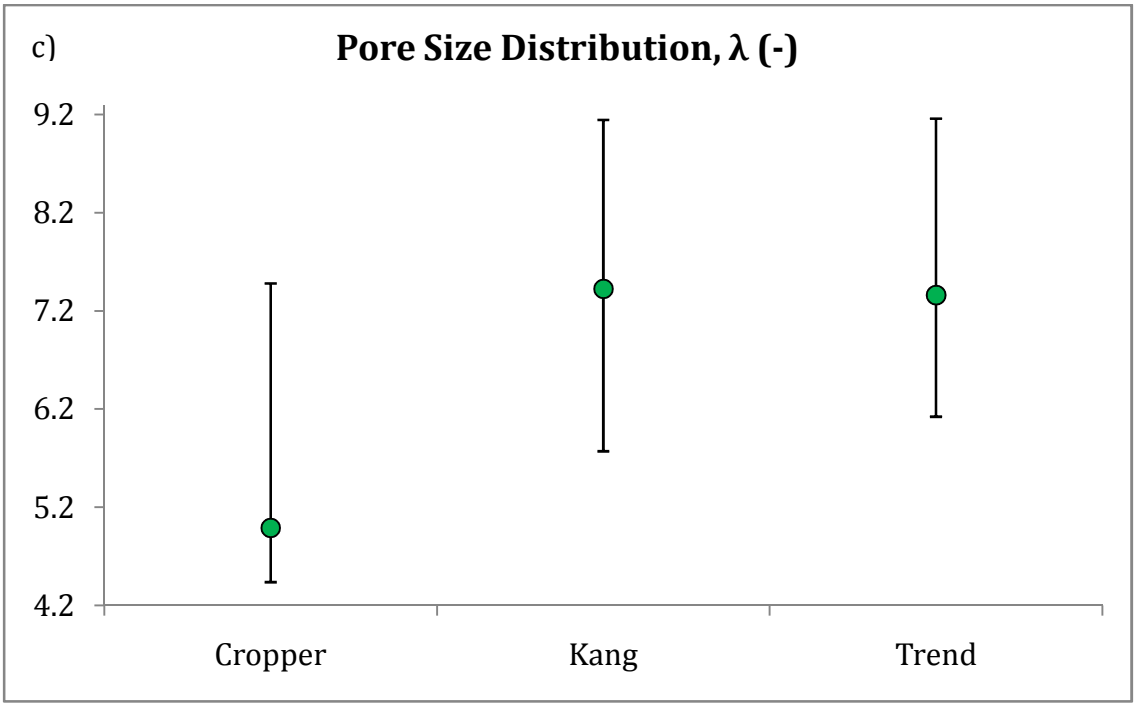


Figure 17 continued



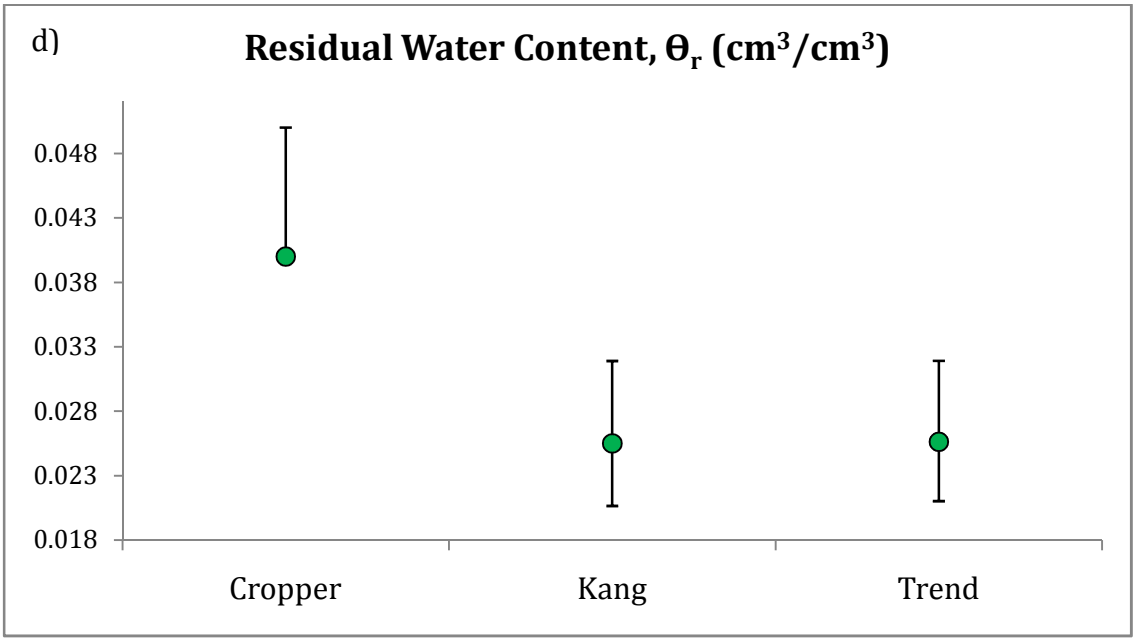


Figure 17 continued

VITA

William “Blake” Roberts was born in Murphy, North Carolina, to William and Tammy Roberts; he has two younger brothers, Baron and Chance, and a younger sister, Kenda. He attended Ranger Elementary School from kindergarten through the eighth grade, and graduated from Hiwassee Dam High school in Murphy, NC in June of 2008. After graduation Blake attended Young Harris College, receiving his Associates of Science Degree in General Science. After graduating from Young Harris College in the May of 2010, Blake attended Western Carolina University where he graduated (in December 2012) Magna Cum Laude with a Bachelor’s of Science degree in geology, concentrating in environmental hydrology, and minoring in physics. He then moved to Knoxville, Tennessee in June of 2013 to begin graduate studies at the University of Tennessee, Knoxville. Blake’s thesis project involved applying geostatistical methods to characterize sub-core scale heterogeneity and identify spatial structure in previously published water retention data obtained using neutron radiography. He also developed a novel trend correction procedure to remove non-geologic heterogeneity and spatial structure thought to be caused by the experimental methods employed. Blake will be graduating with a Master’s of Science in geology in the July of 2015 and plans to enter the oil and gas or environmental consulting fields.