

AN ABSTRACT OF THE THESIS OF

Max Boath for the degree of Master of Science in Sustainable Forest Management presented on November 28, 2018

Title: Estimation of Soil Particle Size Distribution using CT Scans

Abstract approved: _____

Bogdan M. Strimbu

ABSTRACT

Remote sensing techniques have been applied extensively in geospatial investigations, but their use in measuring soil physical attributes has been far less explored. Soil particle size distributions (PSD) are indispensable in modeling pedological and hydrological processes as well as biodiversity. However, estimation of PSD via gravimetric measurement methods, the standard currently in use, is relatively laborious and time-consuming. This research is a pioneering attempt to estimate soil PSD from computerized tomographic (CT) scans. CT scanners non-invasively penetrate three-dimensional (3D) objects to produce a series of two-dimensional (2D) gray images, where grayscale values express density of internal matter in Hounsfield Units (HU). In this study, a model is developed that associates particle size with HU-derived pixels by first classifying the image with an unsupervised technique and then by hierarchically clustering the classes according to soil horizons. The soil PSD is computed as the relative class frequency of classified pixels. For the type of soil used in this study, Weibull distribution was found to fit all layers at a fine 10 mm scale, but a broader horizon-level analysis found lognormal distribution to perform best (in the absence of Weibull). The PSD estimated from CT scans was insignificantly different from the sieve-analysis measured PSD in each horizon. This novel approach to soil diagnostics could transform future soil particle analyses.

©Copyright by Max Boath

November 28, 2018

All Rights Reserved

Estimation of Soil Particle Size Distribution using CT Scans

by
Max Boath

A THESIS

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented November 28, 2018
Commencement June 2019

Master of Science thesis of Max Both presented on November 28, 2018

APPROVED:

Major Professor, representing Sustainable Forest Management

Head of the Department of Forest Engineering, Resources, and Management

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Max Both, Author

ACKNOWLEDGEMENTS

The author expresses sincere appreciation to Professor Bogdan Strimbu for his strong academic direction, personal guidance, and persistent motivation. Without this, I could not imagine a more fruitful, interesting, or enjoyable Masters experience, of which I am grateful to have. Thank you to my parents, Tish and David Boath, for continually supporting my path to higher education, and for the boundless lifelong love and attention. Thanks as well to Rong Fang for the personal encouragement and backing, as well as much appreciated statistical counseling. Additional conceptual help of thesis concepts came from Dr. Sarah Nemanic of the Carlson College of Veterinary Medicine at OSU, and Professor Jeff Hatten of the Department of Forest Engineering, Resources and Management.

TABLE OF CONTENTS

	<u>Page</u>
1 Introduction	1
2 Materials and Methods	3
2.1 Materials.....	3
2.2 Image Acquisition and Pre-Processing.....	4
2.2.1 CT Scanning.....	4
2.2.2 Image Pre-Processing.....	5
2.2.3 Granulometry	8
2.3 Image Processing.....	8
2.4 Fitting Procedure.....	10
3 Results	15
3.1 Soil PSD from Sieve Analysis.....	15
3.2 Image Classification.....	19
3.3 Model Fitting and Comparison	25
4 Discussion.....	30
5 Conclusion	34
References.....	35

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
FIGURE 1. BEASLEY LAKE, MISSISSIPPI, FROM WHICH THE SOIL CORE WAS EXTRACTED.....	4
FIGURE 2. CT IMAGES OF THE SOIL CORE. LEFT: ORIGINAL CT SCAN IMAGES 1-50 (IMAGE 0 NOT PICTURED), CROPPED TO 300 X 1624 PX. SOIL COLUMN WIDTH TAPERS OFF AT THE BEGINNING AND END OF THE SERIES, DUE TO THE CYLINDRICAL NATURE OF THE CORE. RIGHT: ORIGINAL CT RECONSTRUCTION MEASURING 840 MM IN HEIGHT AND 99.3 MM IN DIAMETER.	6
FIGURE 3. CROPPED ORIGINAL CT SCAN IMAGES 18-34, CONCATENATED TO SHOW PROGRESSION OF VERTICAL CROSS SECTION SCANS (SPACING BETWEEN IMAGES NOT TO SCALE.)	7
FIGURE 4. WORKFLOW FOR ESTIMATION OF THE SOIL PSD FROM CT SCANS AND GRANULOMETRY.	14
FIGURE 5A. HISTOGRAMS OF PARTICLE SIZE DISTRIBUTIONS (PSD) FOR SUCCESSFULLY SIEVED 10 MM DEEP CYLINDERS, CM 1-48.....	17
FIGURE 5B. HISTOGRAMS OF PARTICLE SIZE DISTRIBUTIONS (PSD) FOR SUCCESSFULLY SIEVED 10 MM DEEP CYLINDERS, CM 49-84.....	18
FIGURE 6. ISO CLUSTER UNSUPERVISED CLASSIFICATION OF THE 3-BAND IMAGE OBTAINED FROM THE CT SCAN #28. DEPENDING ON THE PARAMETERS OF THE CLASSIFICATION, THE RESULTING IMAGES HAD 7, 8, AND 9 CLASSES. .	19
FIGURE 7. IMAGES 18 – 34, ORIGINAL GRAYSCALE IMAGES (LEFT), AND CORRESPONDING POST-CLASSIFICATION OUTPUTS (RIGHT): 7 CLASSES WERE DETECTED FOR 9 IMAGES AND 8 CLASSES WERE IDENTIFIED FOR 8 IMAGES.	21
FIGURE 8. VISUAL ASSESSMENT OF UNSUPERVISED CLASSIFICATION PERFORMANCE. DISTINCT IMAGE FEATURES IN THE GRAYSCALE INPUT ARE CLASSIFIED IN THE OUTPUT IMAGE USING SPECTRAL PROPERTIES. CT-DERIVED IMAGES PORTRAY MATERIAL DENSITY FROM LEAST (BLACK) TO MOST (WHITE).....	22
FIGURE 9. CLASSIFIED IMAGES 18 - 34, CONCATENATED, VIRTUALLY DEMONSTRATING DIGITAL PORTIONING OF 84 CM DEPTHS VERTICALLY, AS WELL AS HISTOGRAMS FOR SIX SELECTED DEPTHS DISPLAYING AGGREGATED PIXELS-PER-CLASS FROM ALL IMAGES.	24
FIGURE 10. HISTOGRAMS FOR ALL 840 MM DEPTHS. (ONLY 60 OF THE 84 DISTRIBUTIONS WERE EVALUATED TO COMPARE WITH RESPECTIVE GRANULOMETRIC PSD DISTRIBUTIONS, DUE TO ERRORS THAT OCCURRED DURING 24 PSD MEASUREMENTS OF 10-MM TALL DISKS).	25
FIGURE 11. SOIL PARTICLE SIZE DISTRIBUTION OF SIEVE ANALYSIS DATA. FOUR DISTINCT DEPTH-DEPENDENT CLUSTERS ARE APPARENT.	26
FIGURE 12. CLUSTER ANALYSIS ON SOIL PSD DATASETS EXPRESSED AS PERCENTAGES FROM TOTAL: A. SIEVE ANALYSIS B. CLASSIFIED IMAGES.....	27
FIGURE 13. SCREE TEST (THE ROOT MEAN SQUARE DEVIATION IN RED) AND COEFFICIENT OF DETERMINATION, R^2 (IN BLACK), FOR THE CLUSTERS CREATED USING WEIBULL PARAMETERS FROM: A. SIEVE ANALYSIS WEIGHTS, B. CLASSIFIED IMAGES.	28
FIGURE 14. DISTRIBUTIONS OF MEASURED PARTICLE SIZE AND THE PREDICTED DISTRIBUTION FROM CLASSIFIED IMAGES FOR THE FOUR IDENTIFIED HORIZONS, OVERLAID ON THE HISTOGRAM OF THE MEASUREMENTS SUPPLIED BY SIEVE ANALYSIS.	30

LIST OF TABLES

<u>Table</u>	<u>Page</u>
TABLE 1. DISTRIBUTIONS TESTED FOR FITTING THE SOIL PSD OBTAINED BY SIEVE ANALYSIS AND FROM IMAGES. THE A, B, Γ ARE THE PARAMETERS OF THE PROBABILITY DENSITY FUNCTIONS, Γ IS THE GAMMA FUNCTION.	11
TABLE 2. LOGNORMAL DISTRIBUTION PARAMETERS FOR THE FOUR HORIZONS IDENTIFIED FOR THE SIEVE ANALYSIS AND CLASSIFIED IMAGES.	29

1 Introduction

While the chemical composition of soil plays a fundamental role in plants nutrition, the physical properties are primarily responsible for availability of those nutrients. The main numerical measurement describing physical properties of a soil is distribution of particles across horizons, as soils rarely consist of particles of a single size range (Gee and Or, 2002). In essence, particle size distribution (PSD) is a list of ordered values that describes the relative amount (expressed by mass or volume) of granular material present in soil (Gee and Or, 2002; Jillavenkatesa et al., 2001; Ujam and Enebe, 2013). PSD is therefore considered an indication of overall soil texture, as it reveals the composition of particle mass fractions (e.g., clay, silt, sand, gravel) (Bayat et al., 2015; Botula et al., 2013; Gee and Or, 2002; Meskini-Vishkaee and Davatgar, 2018). Soil particles are, in general, convex, irregularly shaped, randomly oriented inorganic matter (Tollner et al., 1998). The arrangement of individual or aggregated particles is often used to predict performance of soil properties, such as the water retention, water capacity, thermal conductivity, even chemical reactivity, as well as land use indications, such as soil degradation, desertification, erosion, and productivity (Arya and Paris, 1981; Bayat et al., 2015; Botula et al., 2013; Fredlund et al., 2002; Gee and Or, 2002; Haverkamp and Parlange, 1986; Hwang and Powers, 2003; Su et al., 2004; Wang et al., 2008). Employing traditional granulometric methods to measure PSDs are laborious and time-consuming (Jillavenkatesa et al., 2001; Nielsen, 2004; Tollner et al., 1998); the commonly used sieve-analysis method is a destructive process that consists of shaking oven-dried soil through a stack of sequentially smaller sieves, then weighing the particles that end up in each size bin (Nielsen, 2004; Tollner et al., 1998; Ujam and Enebe, 2013).

To standardize soil particle size data for consistency across soil types, estimate soil-water characteristic curves, and provide a continuous function over gaps between measured data points, models are often fitted to soil PSDs (Bayat et al., 2015; Botula et al., 2013; Buchan et al., 1993; Esmaeelnejad et al., 2016; Fredlund et al., 2000; Haverkamp and Parlange, 1986; Hwang, 2004; Hwang and Powers, 2003; Zhao et al., 2011). Botula et al. (2013) and Esmaeelnejad et al. (2016) compared various models in capturing PSDs across a regionally diverse collection of soil types, and found that the Weibull model is the most accurate (often "outstanding") fit for soils overall (Bayat et al., 2015; Botula et al., 2013), while additional models (not dismissing Weibull) may

also perform well at fitting particular soil textures types. For example, the lognormal model has been proven to fit coarse-textured soils, such as sandy soils (Botula et al., 2013; Hwang, 2004; Hwang and Powers, 2003; Shirazi and Boersma, 1984), as well as soils with symmetric PSDs (Fredlund et al., 2000, 2002) and with fine-scale grades (Meskini-Vishkaee and Davatgar, 2018). While Esmaeelnejad et al. (2016) argued that models with a higher number of parameters can potentially support better fitting ability, Botula et al. (2013) repeatedly discovered that models with more than three parameters do not necessarily adjust better to the empirical PSD. Nevertheless, models with too many parameters face over-complications, and, therefore, should be avoided due to impracticability (Esmaeelnejad et al., 2016). A multitude of studies focused on modeling PSD suggested that fitting ultimately depends on the grain-size and textural characteristics of the soil. Representing soil PSD as a model aids in the identification and association with other similar soils (Fredlund et al., 2000); this concept of matching models will be applied in this study, by comparing models of a soil's PSD to models of class-frequency statistics derived from remotely sensed imagery.

Remote sensing techniques have been applied extensively to geospatial land cover applications, but their use on underground objects, such as cored soils, has been relatively underexplored. X-ray computed-tomographic (CT) scanners enable remote sensing investigations of solid entities because of the high energy associated with the 0.01 to 10 nanometers wavelength. CT scans non-invasively penetrate a medium from different angles and measure the linear attenuation, which is recorded as a 2D image. The object is subsequently reconstructed by concatenating the image. The final result is a grayscale image, each pixel storing the medium density expressed in Hounsfield Units (HU) (Naveed et al., 2012; Phillips and Lannutti, 1997; Tollner et al., 1998). By convention, darker pixels represent less-dense material (i.e., lower HU), whereas lighter pixels indicate denser material (i.e., higher HU) (Nielsen, 2004).

CT scanners have been used to x-ray soils for at least two decades, but research has primarily focused on measuring soil physical structure, such as pore-space geometry, size, and permeability networks (Beraldo et al., 2014; Elliot et al., 2010; Luo et al., 2010; Mairhofer et al., 2012; Mukunoki et al., 2016; Naveed, 2012; Perez et al., 1999; Peth et al., 2008). Compared to traditional destructive techniques (e.g., sieve analysis), CT scanning offers advantages of time

savings and minimal sample disturbance (Tollner et al., 1998). Despite a steadily growing implementation of CT technology in soil research, its use in predicting soil PSD has been less successful. Tollner et al. (1998) and Nielsen (2004) were similarly motivated to derive PSD from high-quality CT scan images, but the approach of measuring individual grain-sizes in post-filtered imagery was not viable for distinguishing oblique-shaped particles, nor those smaller than the pixel resolution that blend together and, therefore, allow only small samples of soil to be scanned. Moreover, soil CT scan cross-sections do not necessarily split particles at their widest diameters, resulting in measurement errors of individual grains. However, some studies supported the hypothesis that soil density can be used as a proxy to soil PSD. Therefore, the objective of the study is to develop a procedure of expressing accurately and precisely the soil PSD from the CT scans.

2 Materials and Methods

2.1 Materials

A soil core was drilled and collected from beneath Beasley Lake, Sunflower County, Mississippi (33.3982, -90.6762) (Fig. 1). The 25-ha oxbow lake lies in a heavily modified 850-ha alluvial plain watershed in the lower Mississippi River Basin (also known as the Mississippi Delta) (Locke et al., 2005; Wren and Davidson, 2011). Land-use in the region consists mostly of mixed row-crop agricultural (cotton, soybean, rice, as well as catfish), and due to its closed system, runoff into the lake has been the subject of many nutrient, pesticide, and sediment related studies (Cooper et al., 2003; Locke et al., 2005; Wren and Davidson, 2011). The watershed is relatively flat, with maximum change in elevation from the watershed's highest point to lake surface of 5.5 m (Locke et al., 2005); as such, the depth of Beasley Lake is shallow enough (roughly 10 feet or less) that water pressure at the lake bottom would not cause serious sediment compaction. The lake's humid sub-tropic climate has an average annual temperature of 18°C, and produces an annual rainfall of 131 cm (Locke et al., 2005). Soil textures in the watershed vary from sandy loam to heavy clay, with Dundee, Forestdale, Dowling, and Alligator being the major soils represented (Cullum, 2010; Locke et al., 2005).

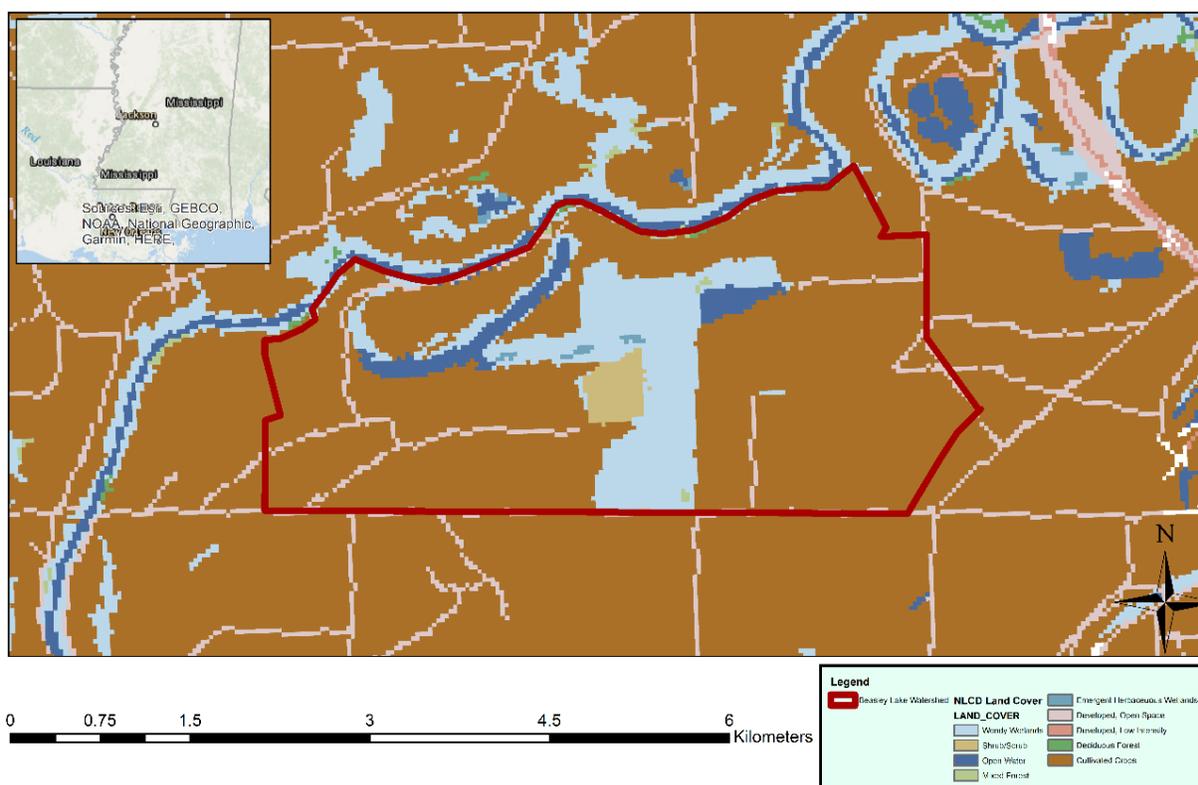


Figure 1. Beasley Lake, Mississippi, from which the soil core was extracted.

The soil sample was cored with an aluminum irrigation pipe, using a vibracorer deployed from a boat, drilled into the bottom of Beasley Lake. The collected vertical sample consisted of maximum internal dimensions measuring 840 mm long, 99.3 mm wide, for a total volume of approximately 6,505 cm³.

2.2 Image Acquisition and Pre-Processing

2.2.1 CT Scanning

A medical-grade Toshiba Aquilion 64-slice Computed Tomography (CT) scanning unit at the Oregon State University College of Veterinary Medicine was utilized to scan the soil core, at a level of 120 kVp and 400 μ A. CT scanners evaluate and translate material density into HU with a linear range (or “window”) from -1000 HU (air) to about +4000 HU (heavy metals) (water being 0 HU) (McGonigal, n.d.). The HU window of the saturated soil sample in this study was presumably 0 (water) to less than 1000 (solid rock, generally less dense than bone). Displayed as a grayscale image on an 8-bit computer monitor, this wide HU window is compressed into

brightness values between 0 and 255 (black:0 to white:255). However, most human eyes can only detect differences in grayscale of every 16 shades of gray (i.e., 6% change) (McGonigal, n.d.) and so a CT scan having a window of 1000 HU translated into 256 brightness values means differences in features might be noticeable at only 60+ HU. In the medical world, changes exist below that threshold; for example, kidneys, pancreas, liver, and blood of humans all register within 25 HU of each other (Heymsfield et al., 2005).

A set comprising 51 concatenated scans captured the vertical cross-sections; the consecutive images being spaced 2 mm apart. The pixel of each image is 0.535898 mm (Figure 2). All images were cropped to 300 x 1624 px ([x, y]) to eliminate wide empty areas outside of the soil column while retaining full spectral detail.

2.2.2 Image Pre-Processing

The cylindrical shape of the core sample lead to narrow images for the first 17 and last 17 scans were deemed unsuitable for ensuing spatial analyses; therefore, they were culled from the image series. This was a matter of needing to overlay a fixed digital window to read all images of the collection, so it was decided that a wider window that captured the full spatial extent of fewer images would yield more accurate results than fitting a narrower window over more images. Considering that such a fixed window could only be as wide as the narrowest image, only the middle 17 images (Images 18-34, the widest 1/3 of the core) were chosen for image processing (Figure 3).

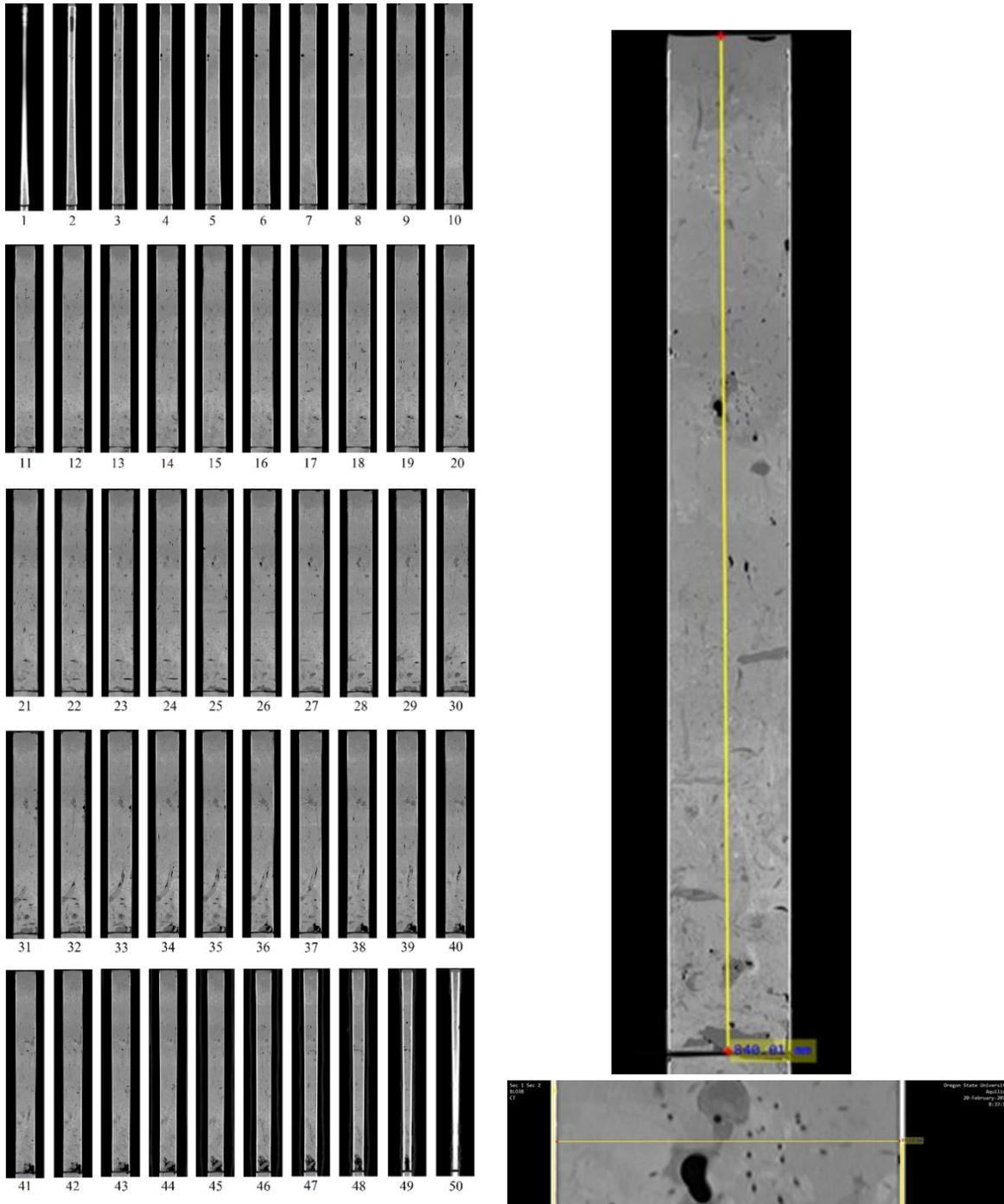


Figure 2. CT images of the soil core. Left: Original CT scan Images 1-50 (Image 0 not pictured), cropped to 300 x 1624 px. Soil column width tapers off at the beginning and end of the series, due to the cylindrical nature of the core. Right: Original CT reconstruction measuring 840 mm in height and 99.3 mm in diameter.

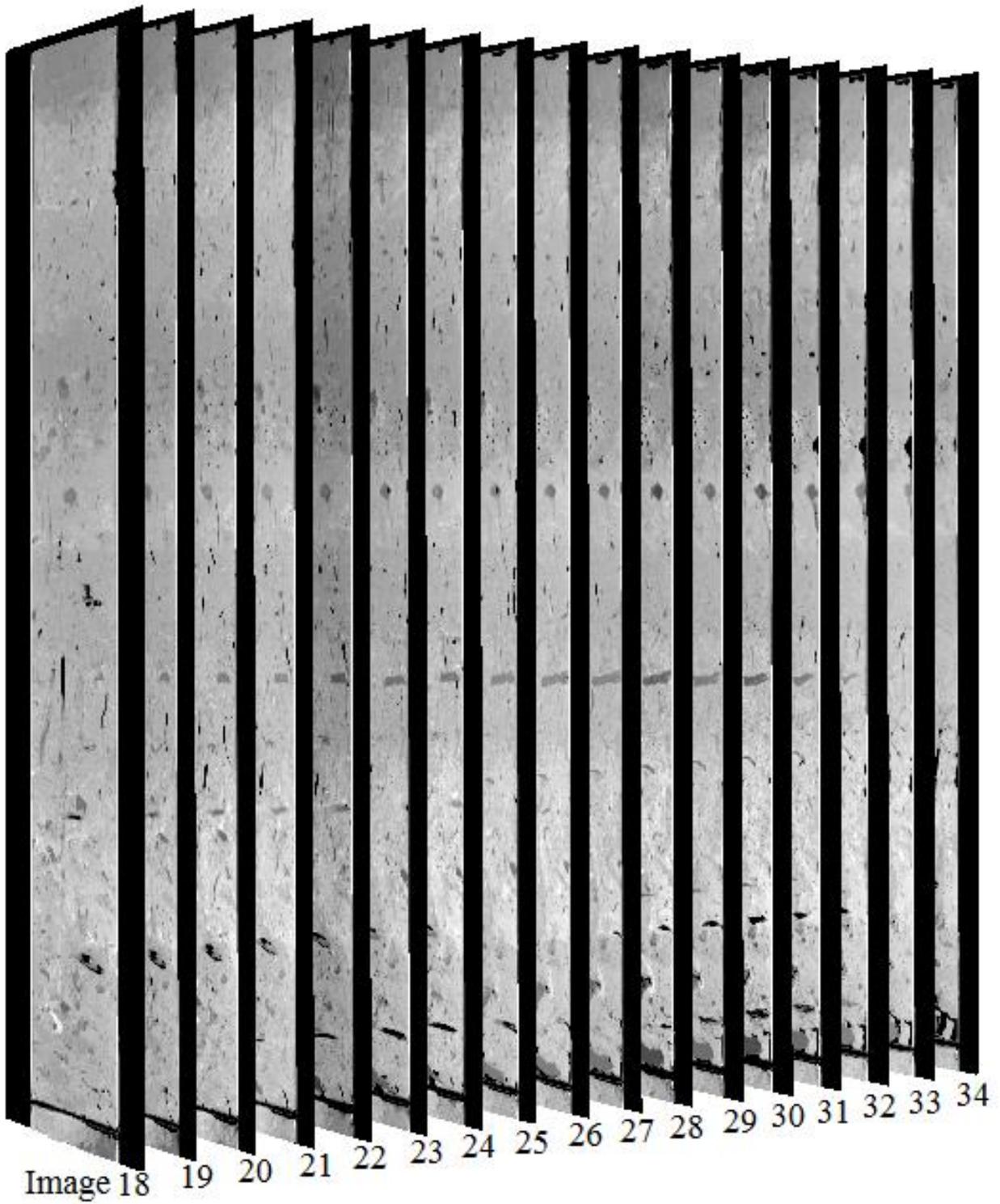


Figure 3. Cropped original CT scan Images 18-34, concatenated to show progression of vertical cross section scans (spacing between images not to scale.)

2.2.3 Granulometry

After being x-rayed, the soil column was manually sliced every 10 mm vertically, which physically divided the core into 84 cm cylinders with diameter 99.3 mm and height 10 mm. Each 10 mm tall soil disks was oven-dried then sieve analyzed, to separate particles into up to 100 normalized size bins ranging from 0.011482 μm - 10000 μm . Some mistakes occurred during granulometric data extraction, which rendered PSD measurements of 24 of the 10 mm tall cylinders unsuitable for analysis. The particles from each bin were weighed, and the recorded PSD measurements of the 60 cylinders (i.e., 600 mm depth) were summarized as histograms (Figure 5).

2.3 Image Processing

A key step in image analysis involves allocation of image features into meaningful categories based on the shared inherent properties within in the image (Naveed, 2012). Because the images display HU units for which no relevant prior information was available, I classified the 17 selected soil core images with the ISO Cluster Unsupervised algorithm, as implemented in ArcGIS 10.4 (Environmental Systems Research Institute, 2008). While several classification algorithms and schemes exist, I selected the ISO Cluster Unsupervised Classification as it is suited for quick and relatively accurate identification of classes within the image without prior training by the user. Alternatively, Supervised Classification would be unsuitable for this application because it requires prior knowledge of image features – not available in this case – and the images contain substantial noise inherent to the creation of the HU that would encumber user-supervised training. Another consideration for not engaging a Supervised Classification approach was due to the imperceptible differences in HUs to the human eye, which rendered identification of classes as biased by the operator. Moreover, a successful performance of Unsupervised Classification in this study would lend more favorably to method repeatability for future soil diagnostics via CT scanning.

The ArcGIS 10.4 implementation of the ISO-Cluster algorithm is an enhancement of the ISODATA algorithm that was developed by Ball and Hall (1965), as it combines the original ISODATA with Maximum Likelihood Classification (Environmental Systems Research Institute, 2008). The ArcGIS procedure initially runs the ISODATA (iterative self-organizing) algorithm (Ball and Hall, 1965), which iteratively samples pixel band values to minimize the

Euclidian distances from a user-specified number of class means (Environmental Systems Research Institute, 2008). To avoid creation of computational artefact spectral classes containing only a few pixels, I decided that individual classes shall contain a minimum of 20 pixels (i.e., $\sim 5.7 \text{ mm}^2$). In the eventuality that not enough pixels were assigned to a class, the class would be eliminated and its pixels merged with an alternative class, thereby resulting in fewer classes than requested. Therefore, a larger-than-expected number of classes should be initially chosen, to gauge the number of unique feature types likely present in the image (Environmental Systems Research Institute, 2008). A trial-and-error process can be used to fine-tune the algorithm, such that the pre-set number of classes is obtained. Because soil particles are generally categorized into three sizes (clay, silt, and sand), I considered that at least two classes should be present in each size. However, the inhomogeneous nature of each soil suggests the possibility of an unbalanced distribution of classes among particle sizes, as well as other non-particle artifacts (e.g., air, larger rocks, roots); therefore, I decided to use at least seven classes, to ensure the range of each particle type (e.g., clay, silt, sand) could be captured by at least one class each.

The ISODATA Unsupervised Classification is followed by a parametric Maximum Likelihood Classification, which assumes that cells are normally distributed around their means. The Maximum Likelihood algorithm uses the covariance of each class's distribution to determine the probability of each pixel to belong to a class (Environmental Systems Research Institute, 2008). Consequently, a pixel may lie closer to one ISO Cluster class mean in spectral space but ultimately is grouped into a different cluster of a more distant mean, due to a higher probability of belonging to a farther mean.

The ISO Cluster Unsupervised Classification requires at least two bands (Environmental Systems Research Institute, 2008); however, CT scanners produce grayscale images of only one band. To accommodate the minimum number of bands requirement and to provide consistency with the vast majority of image classification studies, I created 3-band images that contained the original one-band image triplicated and overlain upon itself. Considering that all pixels will have the same value as a result, collinearity issues could arise that would impact the reliability of the computations. To avoid computational issues, I created two additional 3-band images for all 17 images, which contained along with the original CT scan (band 1), two altered images (band 2 - 3). The altered images were created by adding noise to each HU pixel. I assumed that the noise

should follow a normal distribution with mean 0, to mirror the majority of the procedures used for synthetic data generation. One set of the altered images had variance of noise +/- 1 and +/- 2, and a second set of images with the variance +/- 2 and +/- 4, which added a random variable with a normal distribution and variance σ^2 , $N(0, \sigma^2)$, to the HU at each pixel. The variance size is minute in comparison with the magnitude of the grayscale-value HUs (i.e., < 1%), thereby not significantly altering the information stored in the images. Therefore, for estimating the soil PSD using CT scans I used three sets of 3-band images: 1) a set with three identical bands (i.e., the original CT scans); 2) a set with one band as the original CT scan, a second band as the CT scan plus noise $N(0, 1)$, and a third band as the CT scan plus noise $N(0, 2)$; and 3) a set with one band as the CT scan, a second band as the CT scan plus noise $N(0, 2)$, and a third band as the CT scan plus noise $N(0, 4)$. Considering the particle sizes on which a soil PSD are commonly separated, I classified the images using eight classes, which ensures that at least two classes are present in each broad type of particle size: clay, silt, and sand.

2.4 Fitting Procedure

Identification of a distribution that fits the soil PSD focused on a significant amount of research (Botula et al., 2013; Esmaelnejad et al., 2016). Several distributions were proven to fit most of the soil PST, chief among them being Weibull distribution (Esmaeelenjad). However, other distributions, such as Rossin Rammler, log-normal, or Gompertz, provide superior fit to some soil PSDs, sometimes undistinguishable from Weibull. To be able to estimate soil texture from images, a distribution suited to represent soil PSD must accurately fit different portions of the soils dominated by differing particle size majorities. Therefore, the same distribution should fit all 10 mm cylinders at each of the 84 depths. To mirror the approach presented in many studies aimed at identification of soil PSD (Bayat et al., 2015; Esmaelnejad et al., 2016), ten distributions were tested using the Shapiro-Wilk test: Rossin - Rammler, beta, gamma, Gumbel, lognormal, Pareto, power, Johnson S_b , Jonson S_u , and the 3-parameter Weibull. Model fitting was performed using SAS ver. 9.4 (*SAS*, n.d.).

Table 1. Distributions tested for fitting the soil PSD obtained by sieve analysis and from images. The α , β , γ are the parameters of the probability density functions, Γ is the gamma function.

No.	Distribution	Probability density function
1.	Rosin - Rammler	$e^{-\left(\frac{d}{d_0}\right)^n}$
2.	Beta	$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} x^{\alpha-1}(1 - x)^{\beta-1}$
3.	Gamma	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
4.	Gumbel	$\frac{1}{\beta} e^{-(z+e^{-z})}$
5.	Lognormal	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$
6.	Pareto	$\frac{\alpha x^{\frac{\alpha}{m}}}{x^{\alpha+1}}$
7.	Power	$p\phi(x)(\Phi(-x))^{p-1}$
8.	Johnson S _b	$\gamma + \delta \log\left(\frac{x - \varepsilon}{\varepsilon + \lambda - x}\right)$
9.	Johnson S _u	$\frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1 + \left(\frac{x - \varepsilon}{\lambda}\right)^2}} e^{-\frac{1}{2}(\gamma + \delta \sinh^{-1}\left(\frac{x - \varepsilon}{\lambda}\right))^2}$
10.	Weibull	$\alpha\beta^{-\alpha}(x - \theta)^{\alpha-1} e^{-((x-\theta)/\beta)^{\alpha-1}}$

The same ten models were recast over the aggregated class frequency distributions derived from the classified CT scans, fitting the distribution counts of pixels-per-class. Because the classification of the CT scans was performed for the entire core, whereas the sieve analysis was carried out for individual 10 mm tall cylinders, the same distribution must be present in all images. Otherwise, a mixture of distributions will be present, which would complicate the computations unnecessarily. Therefore, for the ten tested distributions, an additional requirement was imposed on top of the necessity for the model to fit the soil PSD or pixels-per-class data,

namely that the distribution must be the same one to work for all images and for all 10 mm soil disks.

The classification performance of the 3-band images can vary from one CT scan to the next. To ensure that comparisons could be made to classes derived from different CT scanner images, each class must contain pixels with similar HUs; otherwise, particles with different densities will be grouped together, which will invalidate a subsequent separation in soil particle sizes. Considering that for each CT scan the classes will have different ranges of HUs, the constraints that similar HUs should be present in each class translates not to the values themselves but to their continuity. Therefore, each classified CT scan should contain pixels with consecutive HUs values, even if the range of values differs from image to image.

The CT scans provide a description of the soil along the core vertically, whereas the sieve analysis depicts soil granulometry transversally, across the core from top to bottom. To accommodate the two different perspectives (i.e., along and across), the classified images were partitioned into 10 mm sections, similar to the 10 mm tall cylinders. To mirror the fact that the soil cylinder contains information at a particular depth, all the 10 mm sections from all images that were located at the same depth as in the cylinder were merged. For each 10 mm merging (i.e., a set of 17 rectangles), the percentage of each class in respect to all the classes was computed (Eq. 1). The percentage of pixels of a class corresponds to the percentage of a particular particle size, as I assumed that different particle sizes have different densities, therefore different HUs.

$$p_{class\ i} = \frac{1}{n_{images}} \sum_{i=1}^{n_{images}} \frac{\text{number of pixels in class } i}{\sum_{j=1}^{n_{classes}} \text{number of pixels in class } j} \quad 1$$

The wealth of data present in this study, particularly having soil PSD at every 10 mm, is rarely available in real situations. More often, soil PSD is determined not for predefined incremental depths, such as 10 mm, but for entire horizons, such as A_p or B_t (Soil Survey Staff, 1999). To mirror functional reality, the 10 mm tall disks were aggregated according to their PSDs obtained either from sieve analysis or after classification. To ensure realism of the aggregated results, another constraint was imposed that groups must be formed from successive disks, similar to horizons. Because classified images could represent the same particle size with different HUs, predefined classification strategy is unsuitable. Therefore, 10 mm tall disks were

grouped with unsupervised hierarchical clustering, using the average linkage method (Sokal and Michener, 1958) that computes the distance between clusters as:

$$D_{IJ} = \frac{\sum_{i \in I} \sum_{j \in J} \|x - y\|^2}{N_I N_J} \quad 2$$

where D_{IJ} is the distance between clusters I and J ,

N_I and N_J are the number of values in cluster I and J , respectively,

$\|x - y\|^2$ is the Euclidean distance between cluster I and J .

I selected the clusterization method using the average linkage, as it tends to create clusters with similar variances (Massart and Kaufman, 1983). Individual clusters with comparable variances not only represent soil horizons but also balance the groups created from the sieve analysis with the groups created from image classification. I identified significantly different clusters using the scree test (Cattell, 1966) and the procedure proposed by Mojena (1977). I implemented the Mojena procedure, as improved by Milligan and Cooper (1985), which suggested that the number of clusters should be selected based on the change in the distance between two adjacent groups at which:

$$\delta_j = \bar{\delta} + 1.25s_\delta \quad j=1,2,\dots,n \quad 3$$

where $\delta_1, \delta_2, \dots, \delta_n$ are the distance values for stages,

$n, n-1, \dots, 1$ are clusters,

$\bar{\delta}$ and s_δ are the mean and standard deviation of the distances between two adjacent groups, and

$k = 1.25$ is the constant proposed by Milligan and Cooper (1985).

Creation of horizons from images and disks allows development of PSD distributions for each horizon as identified from the sieve analysis and from the CT scans. Nevertheless, the parameters defining the PSD distributions differ according to the source of data (sieve analysis or CT scans). To predict the soil PSD from images, a possible approach is to develop relationships between the parameters defining the two distributions. The PSD distribution of each horizon should be the same as the disks' distribution, as summation of independent random variables with the same distribution is a variable with the same distribution. If the distribution of each

cluster is of the same type irrespective of the source (i.e., sieve analysis or CT scans), relationships between parameters estimated from the images and the parameters estimated from sieve analysis can be developed using seemingly unrelated regression (SUR) (Zellner, 1978):

$$\alpha_{sieve\ analysis} = b_o^\alpha + b_1^\alpha \alpha_{image} + b_2^\alpha \beta_{image} \quad 4$$

$$\beta_{sieve\ analysis} = b_o^\beta + b_1^\beta \alpha_{image} + b_2^\beta \beta_{image} \quad 5$$

According to Kruskal's tree theorem (Kruskal, 1960) the formulation from Eq. 4 and 5 supplies the same results as independent equations estimated with ordinary least squares (OLS). However, I expected that the parameters of the distribution selected for estimation of the soil PSD would not be modeled from the same set of regressors.

The modeling process is summarized in Figure 4.

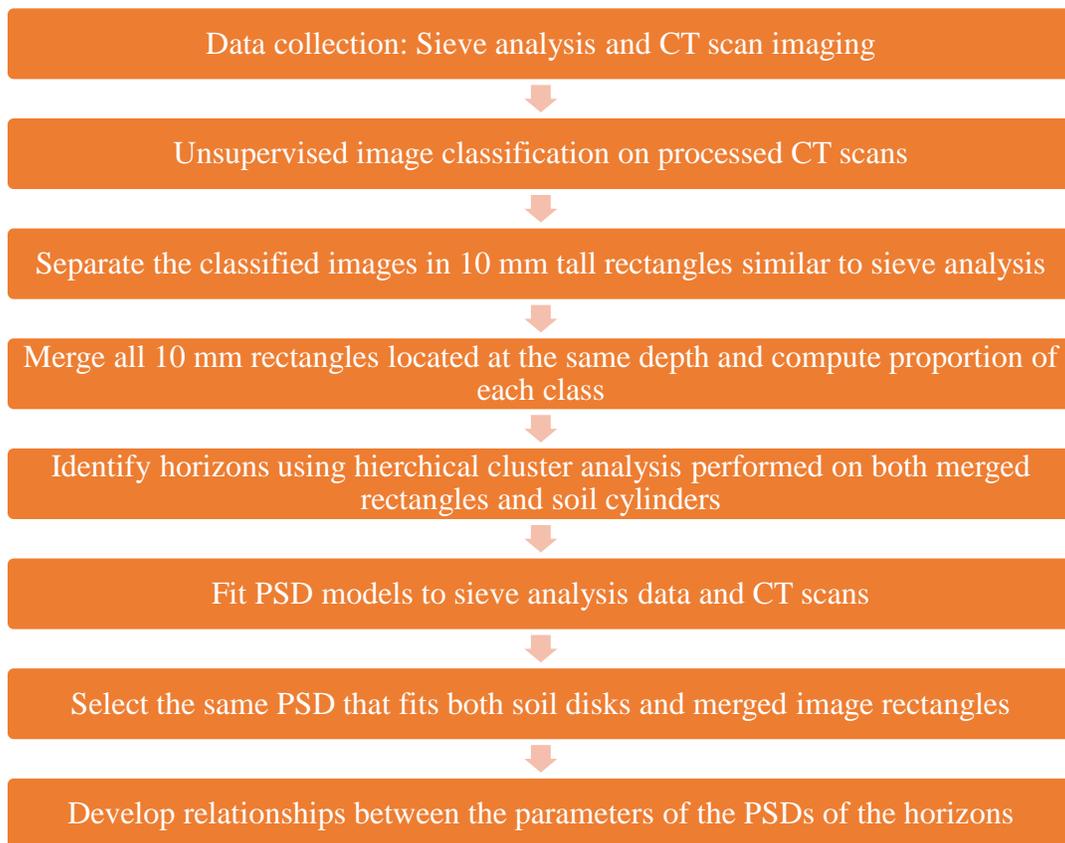


Figure 4. Workflow for estimation of the soil PSD from CT scans and granulometry.

3 Results

3.1 Soil PSD from Sieve Analysis

The histogram created from the sieve analysis for the 10 mm cylinder revealed a skewed bimodal distribution for almost all depths (Figure 5). However, the second mode occurs for particles larger than sand, which not only are of limited interest for plants but also can be estimated quickly and inexpensively. Therefore, for estimating distributions that fit the soil PSD, only the range of particles the size of sand or smaller was used.

Among all ten distributions tested, only Weibull fit all of the 10 mm soil cylinders (p-value > 0.05), followed closely by the lognormal (i.e., almost 90% of the 10 mm cylinder PSDs). The rest of the distributions were either only appropriate for limited portions of the soil core, or faced computational issues. Once the distribution was identified, the parameters were recovered using three procedures: maximum likelihood (ML), as implemented in SAS 9.4; the method of moments; and the percentiles method (Adeyemi, A.A. and Adesoye, P.O., 2016; Hudak and Tiryakioğlu, 2009). In the case of Weibull distribution, the percentiles were computed as

$$q(p, \alpha, \beta) = \beta(-\ln(1 - p))^{1/\alpha} \quad 6$$

and the un-centered moments as

$$m_n = \beta^n \Gamma(1 + \frac{n}{\alpha}) \quad 7$$

where $q(p, \alpha, \beta)$ is the quantile of the probability p and parameters α ,

α and β are the shape and scale parameters of the Weibull distribution

m_n is the un-centered moment of order n .

Therefore, the shape and scale parameters could be estimated as:

$$\alpha = 2/\ln(q_{75}/q_{50})$$

$$\beta = \frac{q_{25}}{(-\ln 0.75)^{(\ln 2)^{-1} \ln(\frac{q_{75}}{q_{50}})}} \quad 9$$

However, several authors have pointed to the sensitivity of outliers when recovering parameters from moments or percentiles, and instead recommend the ML method as the

preferred procedure for parameter estimation. For the soil core data, it was found that the parameters recovered using the method of movements or the percentiles method could vary widely between proximate soil cylinders, suggesting that parameters estimated from ML should be used.

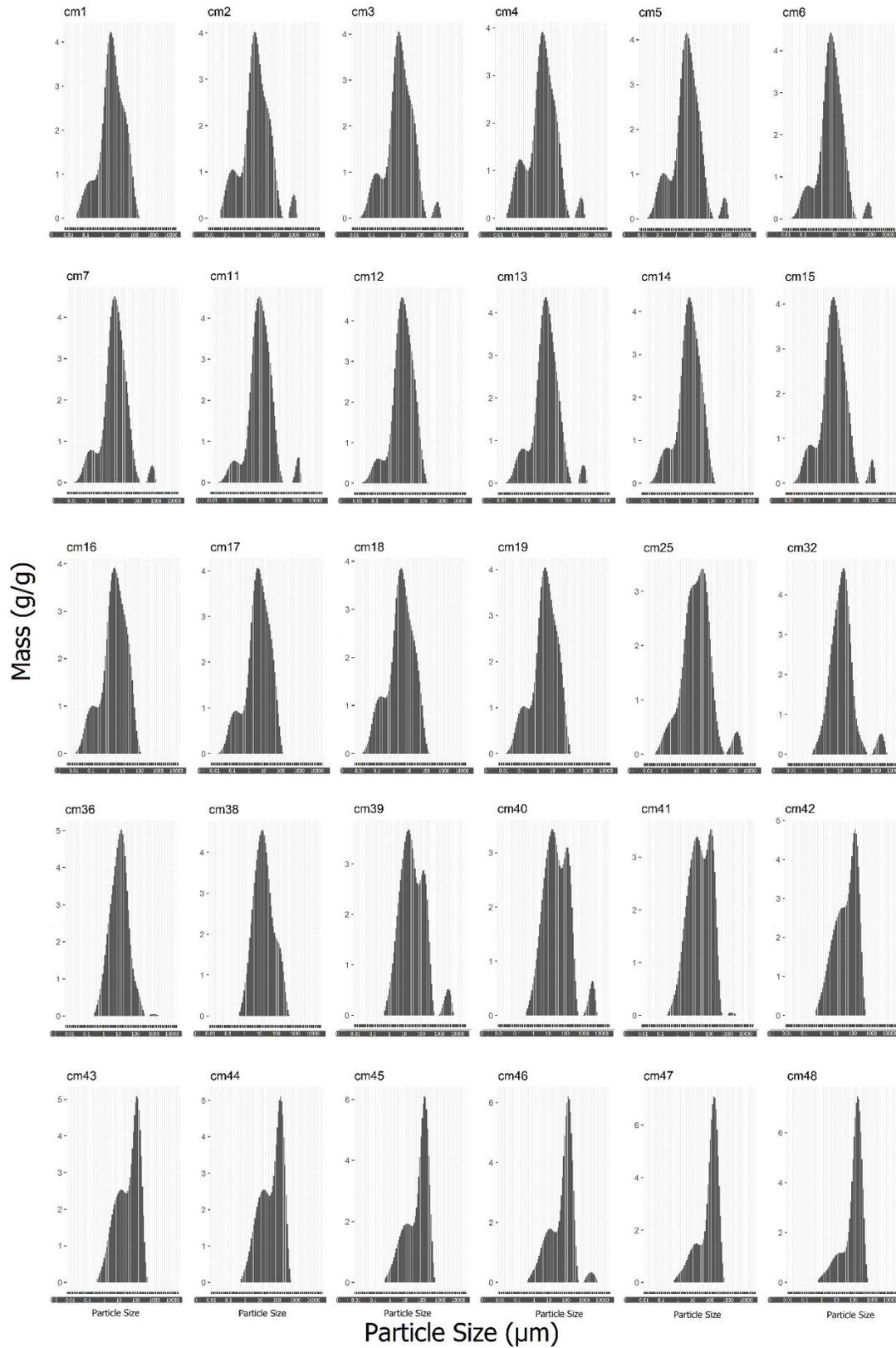


Figure 5a. Histograms of particle size distributions (PSD) for successfully sieved 10 mm deep cylinders, cm 1-48.

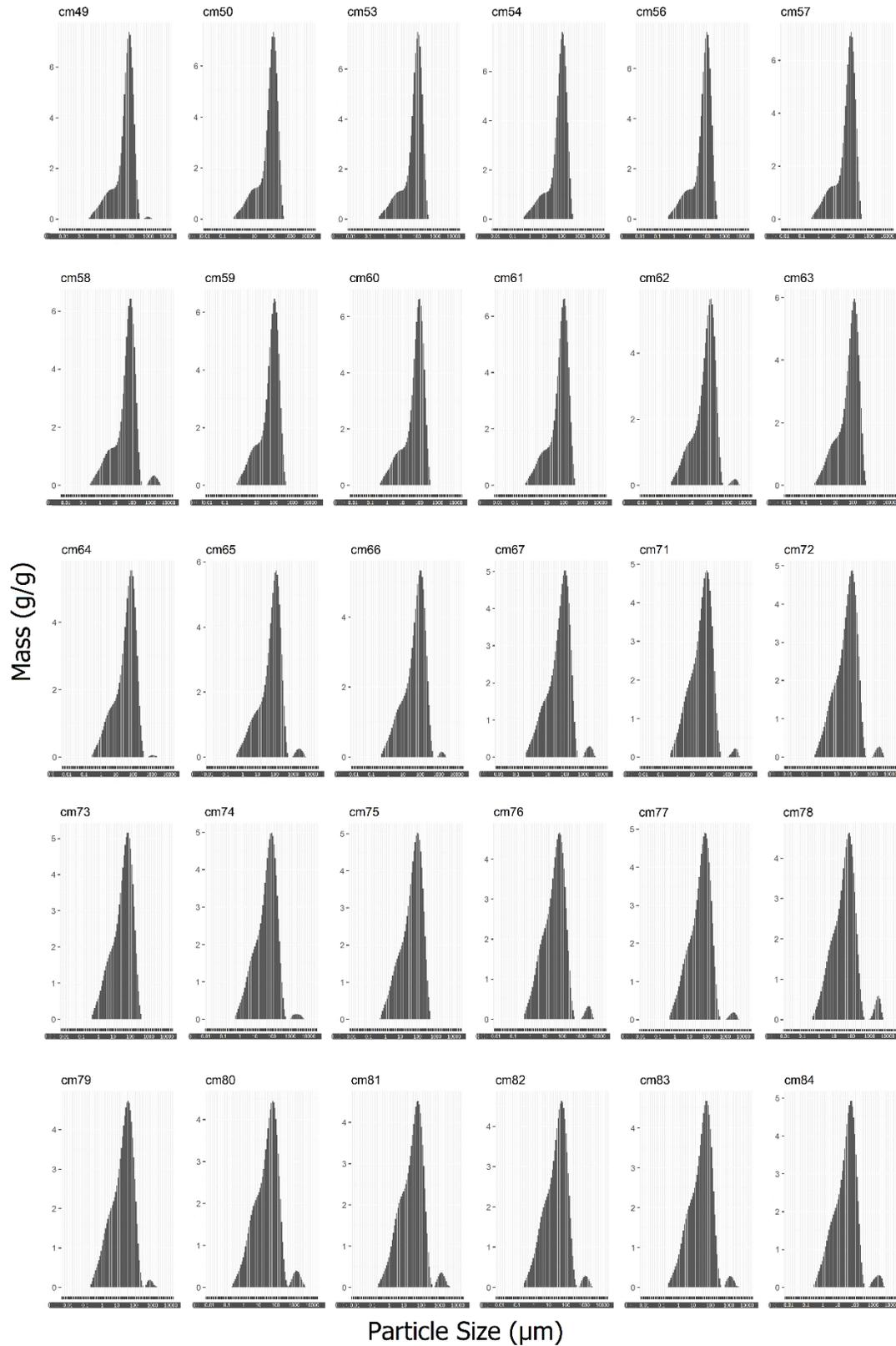


Figure 6b. Histograms of particle size distributions (PSD) for successfully sieved 10 mm deep cylinders, cm 49-84.

3.2 Image Classification

Irrespective of the presence of added variance (noise) or not, all classifications were executed without any computational issues. All classifications started with nine classes and ended with seven, eight, or nine classes (Figure 6), depending on the image. The final number of classes supports the estimation of soil PSD from at least seven classes, which were identified from the images. The lack of computational issues suggested that the analysis should be carried out on the 3-band image with no noise (Figure 6). To allow easier visual distinction of classes both within and among images, all classified images were consistently false-colored (Figure 7).

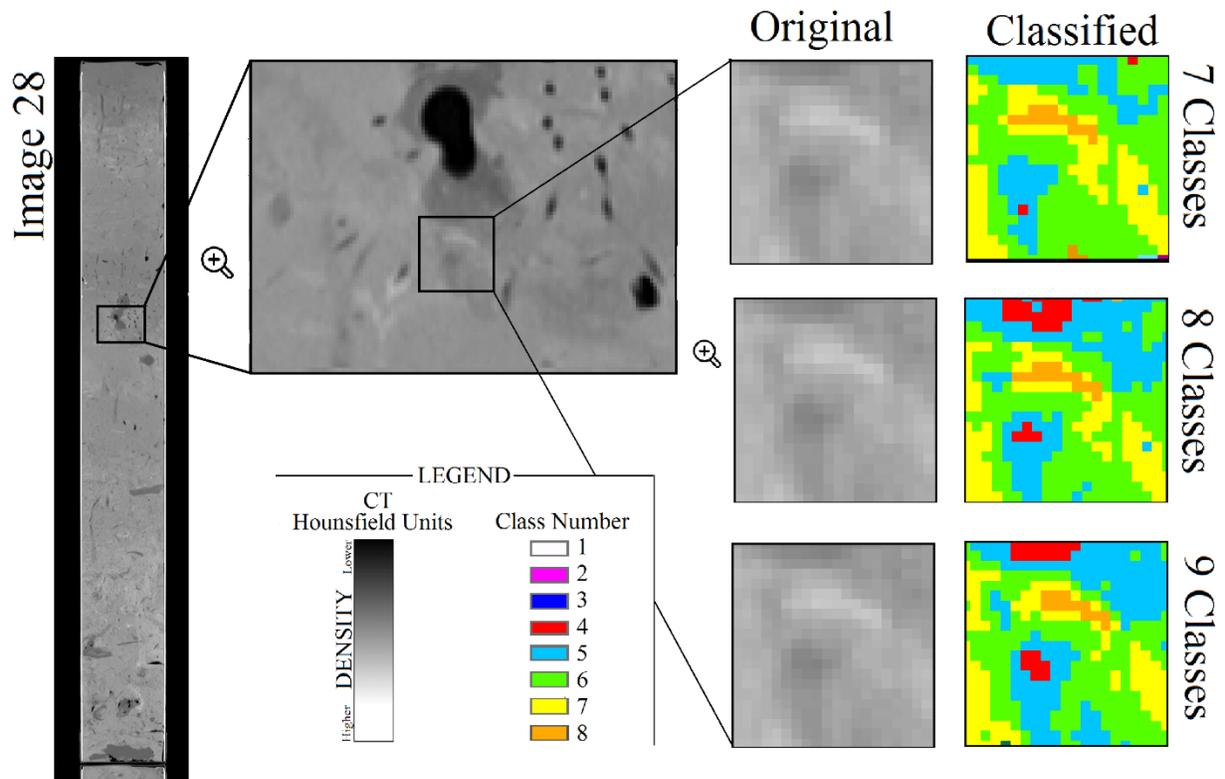


Figure 7. ISO Cluster Unsupervised Classification of the 3-band image obtained from the CT Scan #28. Depending on the parameters of the classification, the resulting images had 7, 8, and 9 classes.

Despite the difference in number of classes among images, the assigned classes generally agreed to the same recognizable features spanning consecutive images. This means that a 7-class image would not necessarily be scaled over classes 1 to 7, but would instead retain sensible consistencies from a different 8-class image; for example, a 7-class image might consist of classes 1 to 3 and 5 to 8, if the usual class 4 feature was too infrequent in the image (fewer than

20 pixels) to constitute its own class (and its pixels then merged into neighboring class identities). I also noticed that some classes (such as classes 1 to 3, or 5 to 6) were sometimes interchangeable between successive images, likely being the same feature in reality, and could perhaps be merged into broader classes signifying definite features (e.g., water, rock, particle size).

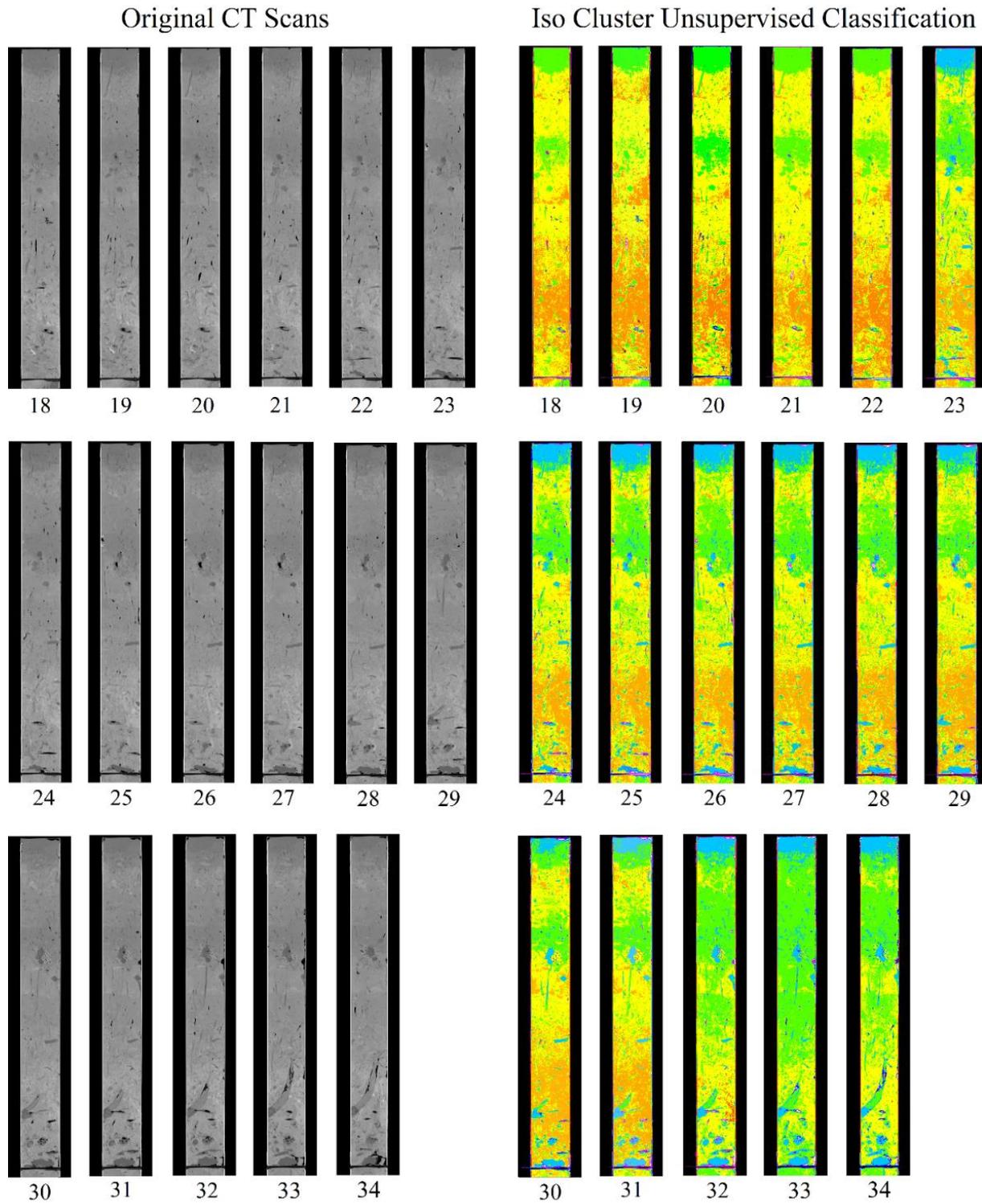


Figure 8. Images 18 – 34, original grayscale images (left), and corresponding post-classification outputs (right): 7 classes were detected for 9 images and 8 classes were identified for 8 images.

The classified images evaluated the identified features appropriately, as distinct observable attributes were present to the pre- and post-classified images (Figure 8). Importantly, class numbers matched detectable grayscale changes apparent in original CT scans. For instance, the algorithm assigned class 1 to the blackest pixels of the original grayscale image, which indicate the least-dense material (e.g., water) detected by the CT scan. Ever-brighter pixels in the grayscale image – which represent increasing material density – aligned with ascending class numbers (up to 8).

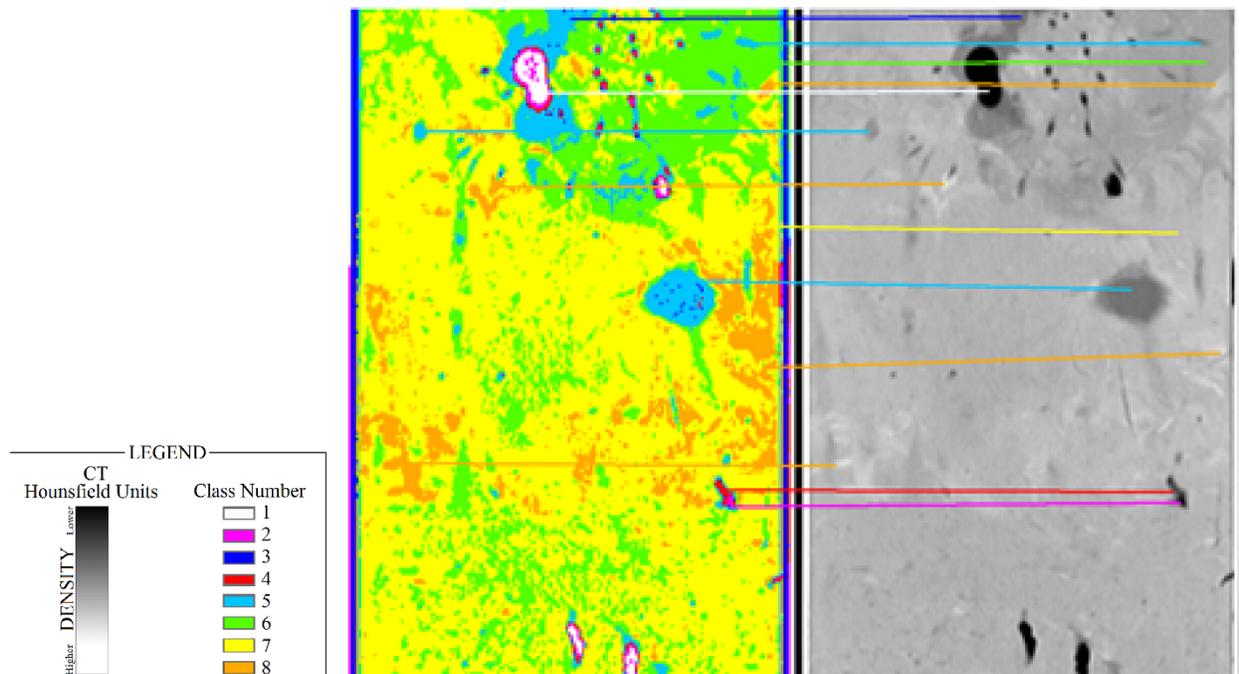


Figure 9. Visual assessment of Unsupervised Classification performance. Distinct image features in the grayscale input are classified in the output image using spectral properties. CT-derived images portray material density from least (black) to most (white).

To compare the distributions of aggregated pixels-per-class from classified images to the distributions of the sieve analysis, each of the 17 classified images was partitioned into 84 equal stacks, representing 10-mm interval depths of the 840 mm soil core. At each centimeter depth (1-84), pixels allocated to each class (1-7 or 1-8) were summed across all images (18-34) (Figure 9). To reflect the proportions of material density present across a graded range I developed histograms of pixels-per-class for each of the 84 depths (Figure 10).

Similar to the sieve analysis, I fit ten distributions to the histograms developed from the classified images. Even though the magnitude of the variables was different (i.e., unitary rather

than sub-unitary), the same findings were achieved: Weibull was the only distribution that accommodated all the depths, followed by lognormal. However, the lognormal distribution did not fit the same depths, which suggested that Weibull distribution should be used to describe soil PSD and consequently to develop linear relationship between the sieve analysis values and the classes from the images. The parameters of the Weibull distribution were also estimated using three methods: ML, method of moments, and the percentiles method. Mirroring the results of sieve analysis, the parameters varied significantly for some nearby pairs of 10 mm depths when estimated with the method of moments and the percentiles method. Therefore, for consistency, I considered only Weibull distribution for modeling, whose parameters were recovered with the ML method.

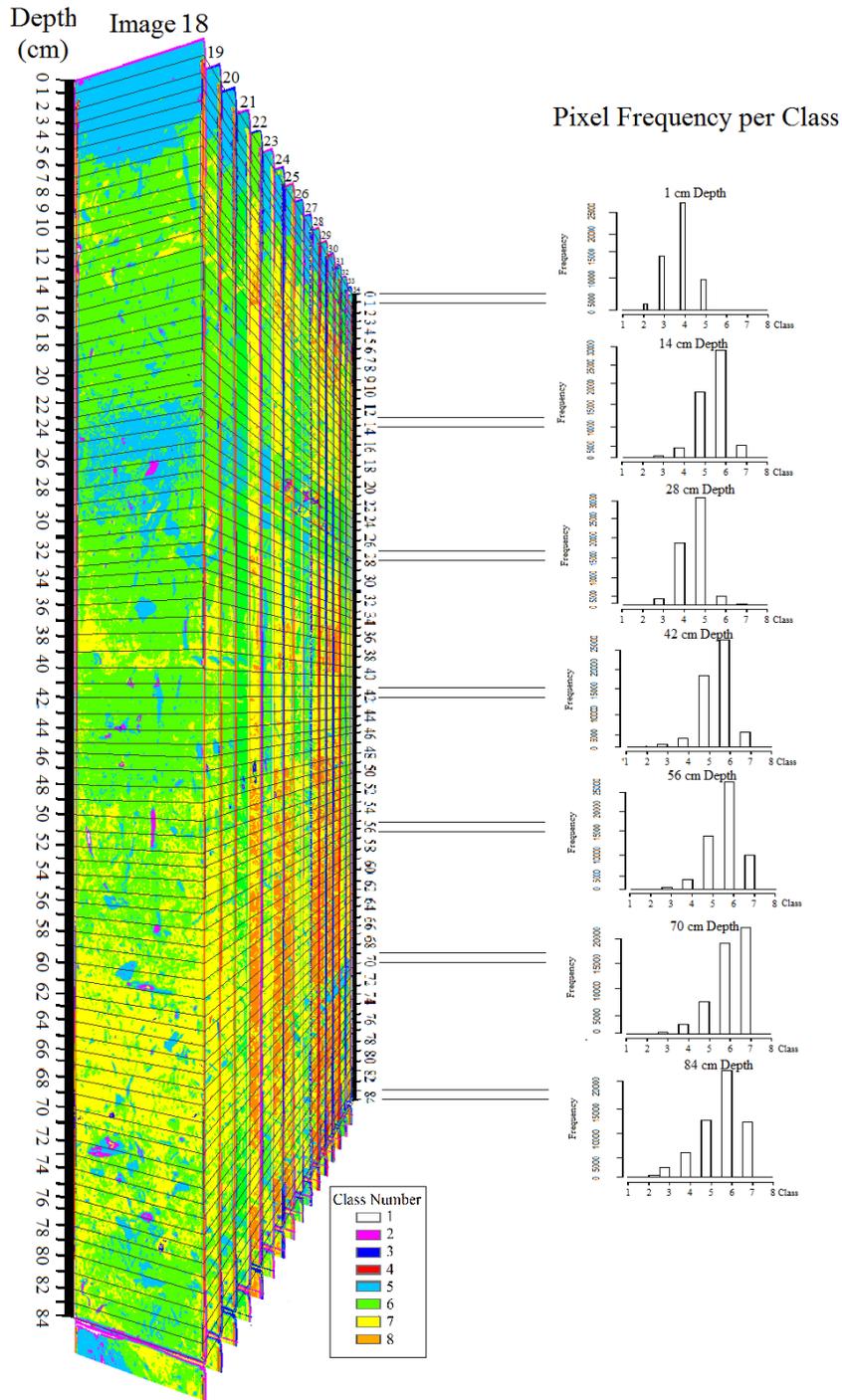


Figure 10. Classified Images 18 - 34, concatenated, virtually demonstrating digital portioning of 84 cm depths vertically, as well as histograms for six selected depths displaying aggregated pixels-per-class from all images.

Histograms Compiled for Images 18–34, at Depths 1–84cm

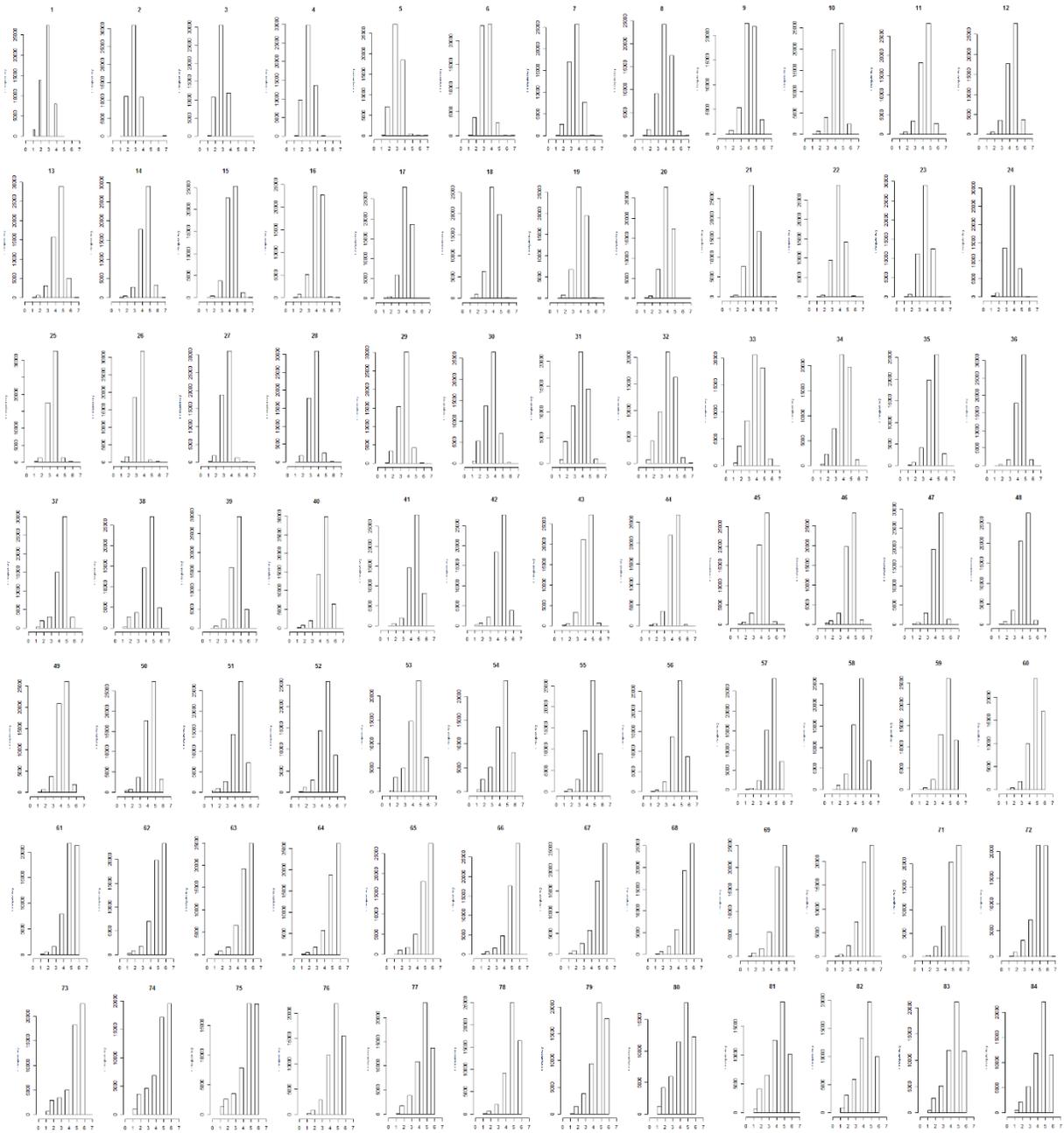


Figure 11. Histograms for all 840 mm depths. (Only 60 of the 84 distributions were evaluated to compare with respective granulometric PSD distributions, due to errors that occurred during 24 PSD measurements of 10-mm tall disks).

3.3 Model Fitting and Comparison

The SUR formalized by Eq. 2 and 4 for the 10 mm tall objects (i.e., cylinders for soil core and set of rectangles for classified images) did not lead to significant relationships between the

parameters of the Weibull distributions estimated from classes and from sieve analysis, which is the objective of the study. The results indicated that the variability along the vertical soil profile is hard to be captured by merging across information. This realization bore the idea of grouping multiple 10 mm cylinders into collective soil horizons. A primary analysis of the effectiveness of this solution was executed by plotting all soil PSDs from sieve analysis (Figure 11), which suggested the presence of at least four unique layers. Despite strong visual potential, execution of a formal analysis would be subsequently needed to verify the true number of layers and to allow for replicating the study.

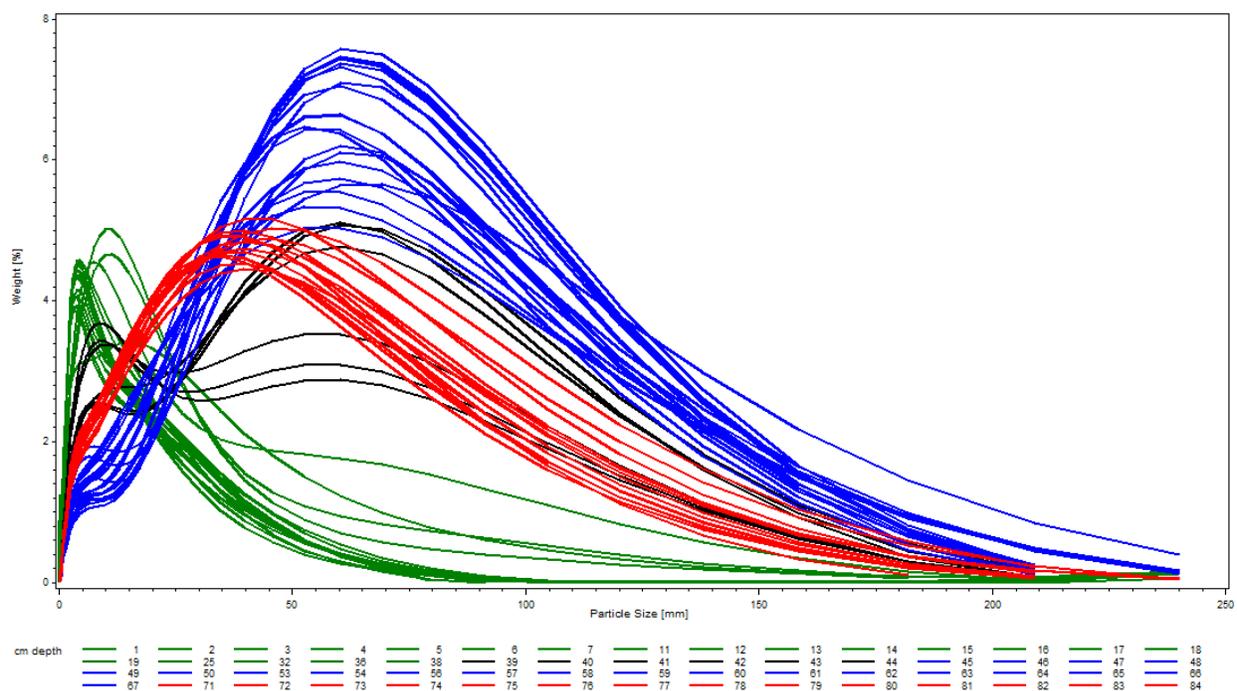


Figure 12. Soil particle size distribution of sieve analysis data. Four distinct depth-dependent clusters are apparent.

To estimate the exact number of soil layers, I executed a hierarchical cluster analysis on both raw data from sieve analysis and image classification (i.e., variables of either class or particle size weighted according to their mass) as well as on the coefficients of the Weibull distribution (i.e., variables alpha and beta). Hierarchical cluster analysis supports the existence of at least four horizons (Figure 12), represented by four clusters, irrespective of the source of data (i.e., sieve analysis or classified images) or level of data-processing (i.e., percentage from total or the Weibull parameters). Almost identical average distances between clusters were obtained when the soil PSD is expressed as weight for each measured particle size, class, or Weibull

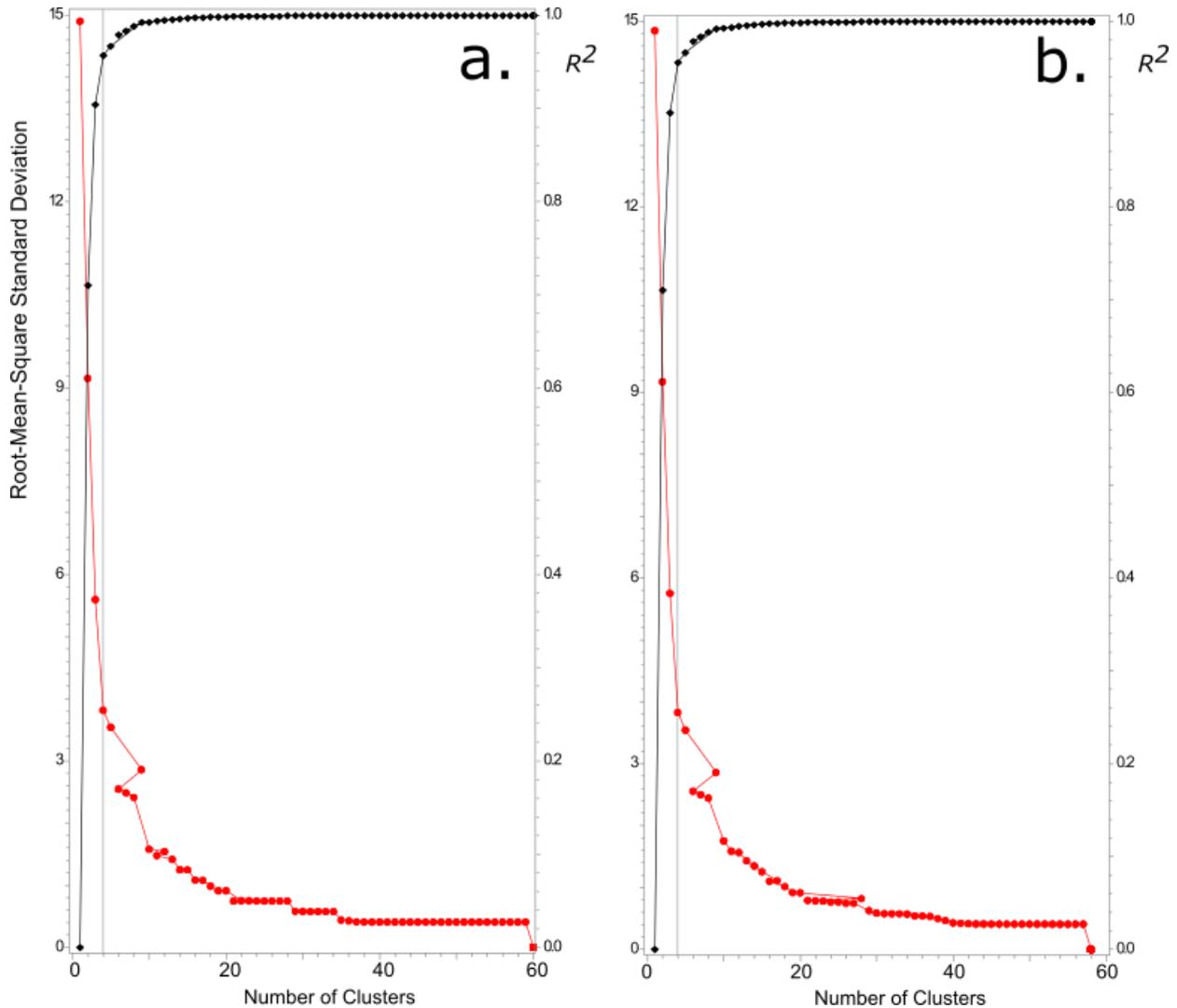


Figure 14. Scree test (the root mean square deviation in red) and coefficient of determination, R^2 (in black), for the clusters created using Weibull parameters from: a. sieve analysis weights, b. classified images.

The horizons identified by cluster analysis served as the basis for estimating soil PSD from CT scans. Since each individual 10 mm cylinder had a Weibull distribution, I decided to use Weibull to similarly describe each horizon. Along with Weibull, I also considered the lognormal distribution, which had been the second most encountered distribution among the 10 mm soil cylinder PSDs. Considering that the sum of independent random variables with different Weibull distributions does not have an analytical form (Nadarajah, 2008), the parameters of the fitted Weibull for the 10 mm cylinders cannot be used for identification of the horizon-level PSD. Consequently, I fit the top two distributions identified for the 10 mm cylinders (i.e., Weibull and lognormal) to each horizon using the ML method. While the Kolmogorov-Smirnov

test indicated that both Weibull and lognormal distribution fit the PSD for both sieve analysis and image classification ($p > 0.03$), the Anderson-Darling test supported only lognormal distribution as fitting all four horizons PSD ($p \geq 0.10$). Therefore, for modeling the soil PSD by horizon I used the lognormal distribution, with parameters presented in Table 2.

Table 2. Lognormal distribution parameters for the four horizons identified for the sieve analysis and classified images.

Horizon	Depth	Scale		Shape	
	[cm]	Sieve Analysis	Image classification	Sieve Analysis	Image classification
1	1 - 20	0.575	1.237	2.413	0.298
2	21 - 51	0.783	1.073	2.293	0.316
3	52 - 69	1.888	0.848	1.655	0.446
4	70 - 84	1.681	0.898	1.774	0.471

The SUR analysis executed on the values from Table 2 revealed that only scale is significant ($p \sim 0.03$), the shape remaining almost constant regardless of horizon ($p > 0.4$). Considering that the same regressors are used for prediction, the parameters of SUR are the same with OLS estimates.

$$scale_{sieve\ analysis} = -1.5 + 6.118 \times shape_{image\ classification} \quad 10$$

$$shape_{sieve\ analysis} = 3.6 - 4.11 \times shape_{image\ classification} \quad 11$$

For the relationships predicting the parameters of the lognormal distribution of the sieve analysis from the parameters of the lognormal distribution of the classified images, the coefficient of determination R^2 was 0.94. The reduced number of observations used for prediction renders most of the tests meaningless, which is the main reason that a p-value of 0.03 was considered acceptable. However, the predicted horizon PSD from the classified images matches the PSD as measured with sieve analysis, with a high degree of confidence ($p\text{-value} > 0.1$) (Figure 14).

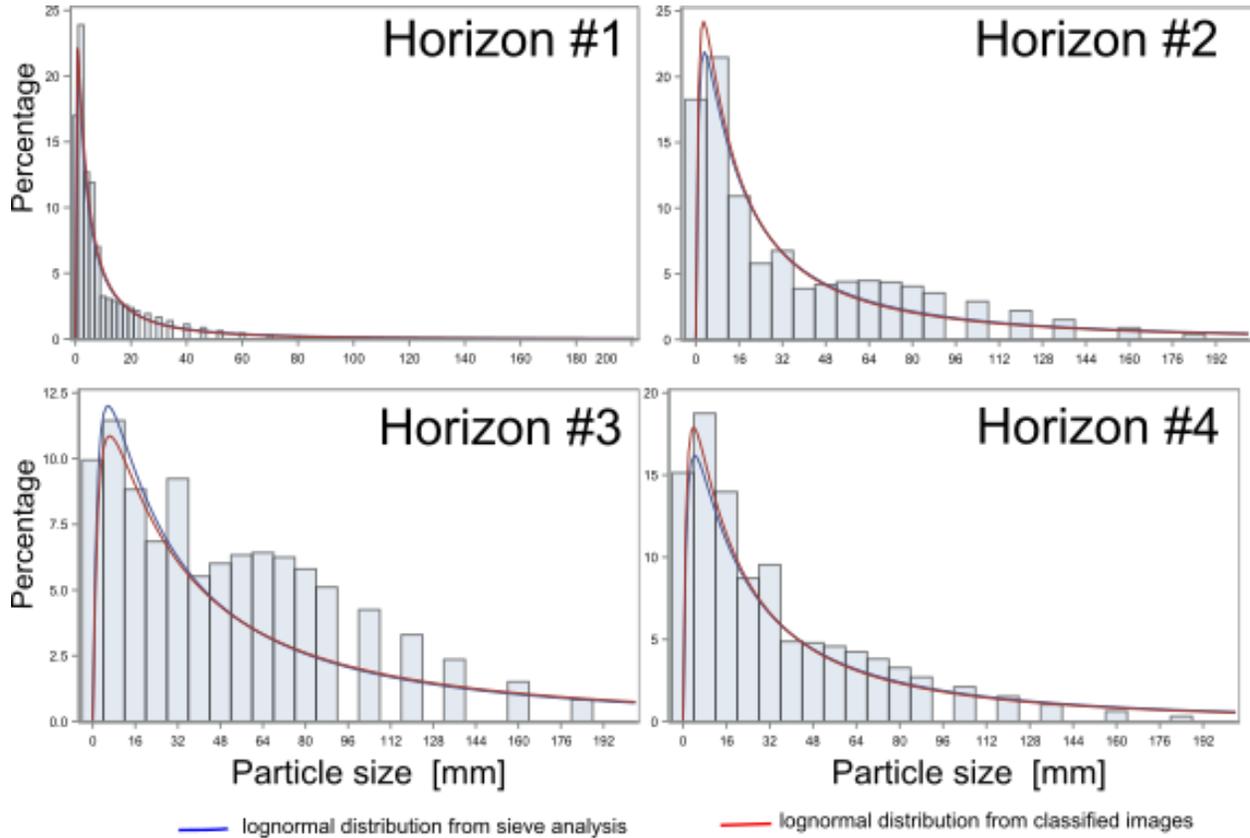


Figure 15. Distributions of measured particle size and the predicted distribution from classified images for the four identified horizons, overlaid on the histogram of the measurements supplied by sieve analysis.

4 Discussion

Results provide statistically significant support for the derivability of precise PSD data from density-imaging scanners, as compared with traditional, arduous, granulometric methods. The study analyzed soil core distributions extracted by hand (sieve analysis) and machine (density-deriving CT scanner). I considered ten models in fitting particle size distributions to determine whether material density can be used as a proxy for predicting PSD. Weibull and lognormal models have previously achieved notoriety for modeling granulometric distributions (Botula et al., 2013; Esmaeelnejad et al., 2016), and these were also initially found to fit the granulometry and imagery datasets best out of the ten tested models: Weibull fit each respective dataset at the individual 10 mm layer level, and lognormal fit the four horizons identified through

clustering. However, Weibull failed to compare ML parameters between both datasets, and ultimately only lognormal showed statistical similarity between them. The successful matching of ML parameters from lognormal was attained more successfully at the broader horizon level rather than individual 10 mm level, which suggests that there is too much variability between individual 10 mm measurements. Such fine-scale data extraction and analysis cannot capture reliable enough data for soil distribution modeling using this method.

Model fitting of CT scan data could potentially be improved by including more of the original CT scan image collection, considering that the dataset in this study was reduced to 1/3 of the total CT scan collection (17 of 51 images). Although the widest, most spatially-adequate subset of images were selected, additional information collected from narrower images might have generated more realistic distributions of image class frequencies, and possibly have increased the tightness of model fits.

Along similar lines, a larger sample size of granulometric PSD measurements could improve comparability with the CT scan-derived dataset, since nearly 30% (24 of 84) of the centimeters in the granulometry dataset had to be forfeited from analysis due to human error. Repetitively measuring soil granulometry from every centimeter depth is tiresome, and time requires constant precision. With many 10 mm soil samples rendered unusable, it is moreover possible that considerable errors persist in the data of the remaining 60 analyzed 10 mm soil disks. The brute dissection of a decimeter-thick soil core dozens of times may have incurred procedural deviations (one considers what happens if a large rock equally bisects a boundary of two depth measurements). Additionally, sieve analysis methods subject soil particles to undesirable disintegration forces (Tollner et al., 1998). Gee and Or (2002) further address the limitations of sieve analysis, specifically that particle size, shape, load, shaking time and motion, and sieve surface geometry all affect the probability of a given particle passing through an opening; therefore reproducibility may only be achieved through careful method standardization. By comparison, digital deconstruction of CT scan imagery is fast, automatable, repeatable without data loss from physical blunders, and likely costs less than other common measurement methods (Tollner et al., 1998). From an efficiency standpoint of time and effort, CT scanning, image cropping, image classification, image statistical computations, and model fitting could all be completed in a single workday. CT scanners can also x-ray wet or waterlogged soil taken

freshly from the ground. In contrast, granulometry requires an extended period of oven-drying (sometimes weeks) to remove all moisture, even before commencing the lengthy process of sieving and weighing.

With a broad-scale (e.g., horizons) fitting method achieving assuring results, the impacts of this method have the potential to be far-reaching. Considering the opportunity for sweeping advancements in PSD diagnostics, promising results from this study should encourage repeated investigation of the methodology onto other soil types. This study explored the viability of using remote sensing methods to analyze just one soil sample of lake-bottom sediment by processing its CT scanned imagery using Iso Cluster Unsupervised Classification. This classification technique, as well as other image segmentation schemes, can partition features captured in any image using the heterogeneity of its spatial and/or spectral thumbprint. Presumably, then, a soil core of any variety or type could be segmented using the same classification and modeling processes trialed in this study. This pioneering study having demonstrated success at larger scales validates the method's ability to estimate soil PSD from CT scanning, and even to potentially predict soil type, since some model distributions have amassed known fitting relationships with particular soils (Esmaelnejad et al., 2016).

Because the soil sample in this study was cored from the bottom of a lake, heterogenous soil horizons were not initially apparent (from preliminary grayscale scans), which obligated the study design to pursue a fine scope of granulometry. However, the hierarchical cluster detection of four horizons, each comprised of 10 mm contiguous cylinders, is backed up through visual investigation of the 17 false-colored classified images that expose consistent divergence of color majorities over four or more layers (Figure 7). This suggests that the CT scanner was not only able to pick up minute detail at nearly particle-level resolution but was also able to detect overall sediment layers, through the changes in density that represent particle size. The four identified horizons could then be modeled more easily than the 60 individual 10 mm layers. This broad horizon-level analysis presumably aligns better with practical applications that are more likely to seek comprehensive PSD of entire horizons or soil types. For example, if one were to study long-term thaw rates of permafrost by scanning soil periodically cored from Tundra to measure receding ice horizon heights and then attempting to correlate between soil type and climate change impact. For such research directed at determining PSD of general horizons, some

granulometric measurements would still be necessary to determine appropriate model parameters and as a check on the CT scan derived PSD model fit, but presumably only one granulometric measurement would be needed for each horizon for which the parameters for that distribution could be used on the distribution of the image data.

Having established a process to estimate soil PSD with significant precision to granulometry, future steps may include experimenting with alternative techniques that aid in classifying soil images with even better precision and more appropriate parameters. CT scanning technology offers many advantages for soil research, but variability in scanning and processing methods nevertheless remains high. Aside from adjustments of CT scanners that can influence remote sensing and image reconstruction (e.g., x-ray attenuation level), several alternative and justifiable image classification schemes exist. For example, the ISO Cluster Unsupervised Classification algorithm was programmed to capture at least 7 classes, and always found either 7 or 8 classes as the optimal amount of distinguishable features per image. In reality, inspection of consecutive classified CT scans suggests that the number of features with like properties might be fewer than the algorithm found (i.e., class numbers 1-3 and 5-6 share the same respective material characterization, which could appropriately be combined into single classes, respectively). This might lead future research involved in advancing our PSD estimation method to experiment with fewer classes and/or meaningful defined categories (i.e., one class each for suspected soil fractions, air, water, and biological). Still, overshooting the number of classes allows for retention of data, better model fitting on more detailed distribution spreads, and better analyst-interpretability of posterior class merging. After all, it is ultimately the role of the user to examine classification results and determine output precision.

In this study, the original grayscale image (band 1) was triplicated into band 2 and band 3, so that the ISO Cluster Unsupervised Classification tool would run. While the outputs of the classification scheme appeared adequate and representative of observable contrast in input grayscale images, adding variance (random noise) to band 2 and/or band 3 might have produced different results and even aided the classification algorithm's ability to tease apart spectral variability inherent within the soil and avoid collinearity issues from a pixel having the same value across all bands. This was not pursued, because there were no computational issues when classifying the no-variance-added 3-band image set.

Alternative software programs that specialize in remote sensing (i.e., ENVI (Harris Geospatial Solutions, 2016)) could also be of use in exploring soil core imagery, their having more extensive spectral classification toolsets and customizable spectral libraries than does ArcGIS currently (i.e., ver. 10.4.1). Finally, CT scanning could become the first step in predicting soil water retention curves (WRC) and soil-water characteristic curves without invasive granulometry, as the methods for determining these heavily-sought metrics currently rely on first obtaining PSD (Arya and Paris, 1981; Fredlund et al., 2002; Haverkamp and Parlange, 1986; Hwang and Powers, 2003). Regardless of the PSD application CT scanning is utilized for, this study outlines the modeling processes that can be used to acquire precise data, remotely.

5 Conclusion

CT scans of soils have previously been subject to similar image segmentation approaches, but were only classified into bimodal images to separate pore space from solid material (Naveed, 2012; Peth et al., 2008). The image classification schemes applied in this study were able to use deconstructed imagery of a soil core to categorize its many physical components into several discrete classes, which could then be connected to meaningful particle sizes by fitting and comparing distribution models against corresponding granulometric PSD.

The use of non-invasive digital technology such as CT scanners holds great potential for advancing real world environmental research and revolutionizing soil diagnostic practices. In this study, the process of predicting physical soil composition was trialed using remote sensing techniques and spatial and statistical analyses. Tested methods delivered results that could successfully diagnose a soil's particle size distribution (PSD) with high precision and efficiency, which could promise to save soil scientists (and others interested in their soils) incredible time and effort in the future. Previous related methods fell short of detecting PSD by relying on scanner image resolution to detect individual particle sizes, but findings from this study demonstrate the ability to confidently, quickly, and non-invasively predict soil PSD through remote sensing of density-derived spectral signatures.

References

- Adeyemi, A.A., Adesoye, P.O., 2016. Evaluation of Some Weibull Parameter Estimation Methods for Characterizing Stem Diameter Distribution in a Tropical Mixed Forest of Southern Nigeria. *Journal of Statistical Science and Application* 4.
<https://doi.org/10.17265/2328-224X/2016.1112.004>
- Arya, L.M., Paris, J.F., 1981. A Physicoempirical Model to Predict the Soil Moisture Characteristic from Particle-Size Distribution and Bulk Density Data. *Soil Science Society of America Journal* 45, 1023.
<https://doi.org/10.2136/sssaj1981.03615995004500060004x>
- Bayat, H., Rastgo, M., Mansouri Zadeh, M., Vereecken, H., 2015. Particle size distribution models, their characteristics and fitting capability. *Journal of Hydrology* 529, 872–889.
<https://doi.org/10.1016/J.JHYDROL.2015.08.067>
- Beraldo, J.M.G., Scannavino Junior, F. de A., Cruvinel, P.E., 2014. Application of x-ray computed tomography in the evaluation of soil porosity in soil management systems. *Engenharia Agrícola* 34, 1162–1174. <https://doi.org/10.1590/S0100-69162014000600012>
- Botula, Y.-D., Cornelis, W.M., Baert, G., Mafuka, P., Van Ranst, E., 2013. Particle size distribution models for soils of the humid tropics. *Journal of Soils and Sediments* 13, 686–698. <https://doi.org/10.1007/s11368-012-0635-5>
- Buchan, G.D., Grewal, K.S., Robson, A.B., 1993. Improved Models of Particle-Size Distribution: An Illustration of Model Comparison Techniques. *Soil Science Society of America Journal* 57. <https://doi.org/10.2136/sssaj1993.03615995005700040004x>
- Cattell, R.B., 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research* 1, 245–276. https://doi.org/10.1207/s15327906mbr0102_10
- Cooper, C., Smith, S., Moore, M., 2003. Surface Water, Ground Water and Sediment Quality in Three Oxbow Lake Watersheds in the Mississippi Delta Agricultural Region: Pesticides. *International Journal of Ecology and Environmental Sciences* 29, 171–184.
- Cullum, R.F., 2010. Effects of Conservation Reserve Program on Runoff and Lake Water Quality in an Oxbow Lake Watershed. *Journal of international environmental application & science* v. 5, 318–328.
- Elliot, T.R., Reynolds, W.D., Heck, R.J., 2010. Use of existing pore models and X-ray computed

- tomography to predict saturated soil hydraulic conductivity. *Geoderma* 156, 133–142.
<https://doi.org/10.1016/j.geoderma.2010.02.010>
- Environmental Systems Research Institute, I., 2008. ArcGIS. ESRI, Redlands CA.
- Esmaelnejad, L., Fatemeh Siavashi, B., Javad Seyedmohammadi, B., Mahmood Shabanpour, B., 2016. The Best Mathematical Models Describing Particle Size Distribution of Soils. Model. Earth Syst. Environ. 2. <https://doi.org/10.1007/s40808-016-0220-9>
- Fredlund, M.D., Fredlund, D.G., Wilson, G.W., 2000. An equation to represent grain-size distribution. *Canadian Geotechnical Journal* 37, 817–827. <https://doi.org/10.1139/t02-080>
- Fredlund, M.D., Wilson, G.W., Fredlund, D.G., 2002. Use of the grain-size distribution for estimation of the soil-water characteristic curve. *Canadian Geotechnical Journal* 39, 1103–1117. <https://doi.org/10.1139/t02-049>
- Gee, G.W., Or, D., 2002. 2.4 Particle-Size Analysis, in: *Methods of Soil Analysis: Part 4 Physical Methods*, SSSA Book Series. Soil Science Society of America, Madison, WI, pp. 255–293. <https://doi.org/10.2136/sssabookser5.4.c12>
- Harris Geospatial Solutions, 2016. ENVI. Exelis Visual Information Solutions, Boulder CO.
- Haverkamp, R., Parlange, J.-Y., 1986. Predicting the water-retention curve from particle-size distribution: 1. Sandy soils without organic matter. *Soil Science* 142.
- Heymsfield, S., Lohman, T., Wang, Z.-M., Going, S., 2005. Human Body Composition, in: *Human Body Composition*. p. Chapter 7.
- Hudak, D., Tiryakioğlu, M., 2009. On estimating percentiles of the Weibull distribution by the linear regression method. *Journal of Materials Science* 44, 1959.
<https://doi.org/10.1007/s10853-009-3306-1>
- Hwang, S.I., 2004. Effect of texture on the performance of soil particle-size distribution models. *Geoderma* 123, 363–371. <https://doi.org/10.1016/j.geoderma.2004.03.003>
- Hwang, S.I., Powers, S.E., 2003. Lognormal Distribution Model for Estimating Soil Water Retention Curves for Sandy Soils. *Soil Science* 168, 156–166.
<https://doi.org/10.1097/01.ss.0000058888.60072.e3>
- Jillavenkatesa, A., Dapkunas, S.J., Lum, L.-S.H., 2001. Particle Size Characterization. National Institute of Standards and Technology.
- Kruskal, J.B., 1960. Well-Quasi-Ordering, The Tree Theorem, and Vazsonyi's Conjecture. *Transactions of the American Mathematical Society* 95, 210–225.

- <https://doi.org/10.2307/1993287>
- Lake, B., n.d. CEAP Project Plan 2/7/05 Management Practices.
- Locke, M.A., Knight, S.S., Cooper, C.M., Smith, S., Moore, M.T., Shields, F.D., Cullum, R.F., 2005. Beasley Lake, Mississippi CEAP Project Plan. USDA NRCS.
- Luo, L., Lin, H., Li, S., 2010. Quantification of 3-D soil macropore networks in different soil types and land uses using computed tomography. *Journal of Hydrology* 393, 53–64. <https://doi.org/10.1016/j.jhydrol.2010.03.031>
- Mairhofer, S., Zappala, S., Tracy, S.R., Sturrock, C., Bennett, M., Mooney, S.J., Pridmore, T., 2012. RooTrak: Automated Recovery of Three-Dimensional Plant Root Architecture in Soil from X-Ray Microcomputed Tomography Images Using Visual Tracking. *Plant Physiology* 158, 561–569. <https://doi.org/10.2307/41435391>
- Massart, D.L., Kaufman, L., 1983. The interpretation of analytical chemical data by the use of cluster analysis. New York: Wiley.
- McGonigal, M., n.d. CT Scan Images Simplified [WWW Document]. The Trauma Professional's Blog. URL <http://regionstraumapro.com/post/16349545265> (accessed 5.7.18).
- Meskini-Vishkaee, F., Davatgar, N., 2018. Evaluation of Different Predictor Models for Detailed Soil Particle-Size Distribution. *Pedosphere* 28, 157–164. [https://doi.org/10.1016/S1002-0160\(17\)60422-3](https://doi.org/10.1016/S1002-0160(17)60422-3)
- Milligan, G.W., Cooper, M.C., 1985. An Examination of Procedures for Determining the Number of Clusters in A Data Set. *Psychometrika* 50, 159–179.
- Mojena, R., 1977. Hierarchical grouping methods and stopping rules: an evaluation. *Comput J* 20, 359–363. <https://doi.org/10.1093/comjnl/20.4.359>
- Mukunoki, T., Miyata, Y., Mikami, K., Shiota, E., 2016. X-ray CT analysis of pore structure in sand. *Solid Earth* 7, 929–942. <https://doi.org/10.5194/se-7-929-2016>
- Nadarajah, S., 2008. A Review of Results on Sums of Random Variables. *Acta Appl Math* 103, 131–140. <https://doi.org/10.1007/s10440-008-9224-4>
- Naveed, M., Moldrup, P., Arthur, E., Wildenschild, D., Eden, M., Lamande, M., Vogel, H.-J., de Jonge, L.W., 2012. Revealing Soil Structure and Functional Macroporosity along a Clay Gradient Using X-ray Computed Tomography. *Soil Science Society of America Journal* 77, 403–411. <https://doi.org/10.2136/sssaj2012.0134>
- Nielsen, B.D., 2004. NON-DESTRUCTIVE SOIL TESTING USING X-RAY COMPUTED

- TOMOGRAPHY. Montana State University.
- Perez, K.T., Davey, E.W., Moore, R.H., Burn, P.R., Rosol, M.S., Cardin, J.A., Johnson, R.L., Kopans, D.N., Kopans2, D.N., 1999. Application of Computer-Aided Tomography (CT) to the Study of Estuarine Benthic Communities. Source: Ecological Applications Ecological Applications 9, 1050–1058.
- Peth, S., Horn, R., Beckmann, F., Donath, T., Fischer, J., Smucker, A.J.M., 2008. Three-Dimensional Quantification of Intra-Aggregate Pore-Space Features using Synchrotron-Radiation-Based Microtomography. Soil Science Society of America Journal 72, 897–907. <https://doi.org/10.2136/sssaj2007.0130>
- Phillips, D.H., Lannutti, J.J., 1997. Measuring physical density with X-ray computed tomography. NDT&E International 30, 339–350.
- SAS, n.d.
- Shirazi, M.A., Boersma, L., 1984. A Unifying Quantitative Analysis of Soil Texture. Soil Science Society of America Journal 48, 142–147.
- Soil Survey Staff, 1999. Soil Taxonomy; A Basic System of Soil Classification for Making and Interpreting Soil Surveys, 2nd ed. USDA NRCS.
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. The University of Kansas Science Bulletin 28, 1409–1438.
- Su, Y.Z., Ha, L.Z., Wen, Z.Z., Zhang, T.H., 2004. Fractal features of soil particle size distribution and the implication for indicating desertification. Geoderma 122, 43–49.
- Tollner, E.W., Melear, N D, Rodriguez, L A, Wright, M E, Melear, Nathan D, Rodriguez, Luis A, Wright, Malcolm E, 1998. Soil aggregate size distributions using x-ray images. Transactions of the ASAE 41, 1207–1215.
- Ujam, A.J., Enebe, 2013. Experimental Analysis of Particle Size Distribution using Electromagnetic Sieve. American Journal of Engineering Research 02, 2320–847.
- Wang, D., Fu, B.-J., Zhao, W., Hu, H., 2008. Multifractal characteristics of soil particle size distribution under different land-use types on the Loess Plateau, China. Catena 72, 29–36.
- Wren, D.G., Davidson, G.R., 2011. Using lake sedimentation rates to quantify the effectiveness of erosion control in watersheds. Journal of Soil and Water Conservation 66, 313–322. <https://doi.org/10.2489/jswc.66.5.313>
- Zellner, A., 1978. Estimation of functions of population means and regression coefficients

including structural coefficients : A minimum expected loss (MELO) approach. *Journal of Econometrics* 8, 127–158.

Zhao, P., Shao, M., Horton, R., 2011. Performance of soil particle-size distribution models for describing deposited soils adjacent to constructed dams in the China Loess Plateau. *Acta Geophysica* 59, 124–138. <https://doi.org/10.2478/s11600-010-0037-2>