AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research®

TECHNICAL REPORTS: METHODS

10.1029/2021WR030095

Soil Classification: A New Approach for Grouping Soils Using Unsaturated Hydraulic Conductivity Data

Behzad Ghanbarian[1] **and Brandon A. Yokeley**[1]

[1]Department of Geology, Porous Media Research Lab, Kansas State University, Manhattan, KS, USA

Key Points:

- We developed a new approach to classify soils based on similarities in their hydraulic properties
- Using concepts from critical path analysis, soils with similar critical pore size at the same effective water saturation were grouped together
- 102 samples from the UNSODA database were analyzed using the proposed approach, and eight soil types were detected

Correspondence to:

B. Ghanbarian,
ghanbarian@ksu.edu

Citation:

Ghanbarian, B., & Yokeley, B. A. (2021). Soil classification: A new approach for grouping soils using unsaturated hydraulic conductivity data. *Water Resources Research*, *57*, e2021WR030095. https://doi.org/10.1029/2021WR030095

Received 29 MAR 2021
Accepted 11 AUG 2021

Abstract Grouping soils based on similarities in their textural, taxonomic, and/or structural properties has broad applications to pedology, hydrology, and soil science. In this study, we present a new approach for classifying soils using hydraulic conductivity data. We apply concepts from critical path analysis and calculate critical pore sizes at various water saturations from the unsaturated hydraulic conductivity curves. Soils with similar critical pore size at the same effective water saturation are then grouped into the same class. To demonstrate the practical application of the proposed soil classification method, we use 102 samples including nine soil textures from the UNSODA database. Applying a curve clustering method, we find eight different soil classes within the studied data set.

1. Introduction

Soil classification is a routine approach for grouping soils based on their intrinsic property, behavior, or genesis with broad applications to soil science, particularly pedometry, soil survey, and soil mapping. In the literature, grouping soils has been performed on the basis of similarities in taxonomic (Devine et al., 2021; Pachepsky & Rawls, 1999), textural (García-Gutiérrez et al., 2018; Pachepsky & Park, 2015), and structural (Nguyen et al., 2015; Rawls & Pachepsky, 2002a) properties. For example, the Natural Resource Conservation Service (NRCS, 1986) classified soils into four hydrologic groups based on the soil's runoff potential. The four Hydrologic Soils Groups are A, B, C, and D where Soil Groups A and D generally have the smallest and greatest runoff potentials, respectively.

Soil classification has been a long-standing challenge that recently became even more important, particularly for global-scale studies where (a) spatial and temporal variations in soil hydraulic properties exceed the field sampling capabilities (Twarakavi et al., 2010) and (b) a huge amount of data collected on a short-time period basis should be analyzed and interpreted (Dennis & Berbery, 2021). In the literature, properties such as textural classes and particle size distributions were widely used to group soils based on their textural similarities (García-Gutiérrez et al., 2018; Pachepsky & Rawls, 1999; Pachepsky et al., 2006; Rawls & Pachepsky, 2002b). For example, Pachepsky and Rawls (1999) investigated if soil grouping based on their taxonomic unit, moisture and temperature regime, and textural class would improve the accuracy and reliability of pedotransfer functions using 447 soil samples from the Oklahoma National Resource Conservation Service database. They applied the Group Method of Data Handling (GMDH) approach to develop regression-based relationships and found that grouping improved the accuracy of pedotransfer functions in most cases. They stated that, "Although PTFs [pedotransfer functions] developed from the groups were more accurate than the PTFs developed from the whole database, they were not more reliable. Improving PTF reliability may be an issue distinctly different from improving PTF accuracy."

Twarakavi et al. (2010), however, questioned the validity of texture-based classification from a hydraulic point of view. Those authors applied the *k*-means clustering method in combination with the ROSETTA pedotransfer functions to determine soil hydraulic classes. They found that the texture-based classification did not group soils satisfactorily when there was a considerable impact of capillary forces.

One of the pioneering approaches in classifying soils is using the Miller-Miller similar-media theory (Miller & Miller, 1956) in which all regions in a soil are assumed to be structurally identical magnifications of a reference location. This theory has been widely used to classify porous media based on their hydraulic properties, such as capillary pressure and hydraulic conductivity curves (Vereecken et al., 2007). In the Miller-Miller theory, two soils are similar when scale factors can be found that transform the behavior of one to that of the other one (Nielsen et al., 1998). Generally, the Miller-Miller theory is valid as long as media are similar

either with respect to their pore space characteristics or their hydraulic properties. Miller and Miller (1956) stated that, "in practice, the occurrence of detailed similarity throughout the microscopic geometries of two media has zero probability." However, there is no exact methodology to explore such a geometric similarity within a microscopic domain (Sadeghi et al., 2016). Accordingly, most studies investigated the similarity of macroscopic and/or measurable characteristics, such as average grain size (Klute & Wilkinson, 1958).

There is experimental evidence showing that rigorous criteria for the Miller-Miller similarity may be held in sand packs (Klute & Wilkinson, 1958; Schroth et al., 1996). Although similarity is required for the validity of the Miller-Miller theory, Sadeghi et al. (2016) recently discussed that similarity is not the only required criterion. They demonstrated that the interrelation between water retention and hydraulic conductivity curves is also essential to scale and classify porous media. They argued that the interrelation depends not only on the pore space geometry, but also on solid-liquid interactions, and, thus, similar interrelation cannot be drawn from the similarity of microscopic pore space geometry. Sadeghi et al. (2016) proposed a joint scaling factor, a parameter that is, a function of air entry value and saturated hydraulic conductivity (see their Table 1), to evaluate the interrelation among 26 soils and found six groups of similar soils.

The portion of the subsurface above the groundwater is the unsaturated zone. Proper investigation of the unsaturated zone requires characterizing hydraulic properties of soils under variably saturated conditions. Therefore, the main objective of this study is to develop a new approach that classifies soils based on similarities in their hydraulic properties and two-phase flow characteristics. Our theoretic methodology is based on concepts from critical path analysis that has been successfully applied to estimate unsaturated hydraulic conductivity from soil water retention curve (Ghanbarian & Hunt, 2017; Ghanbarian, Hunt, & Daigle, 2016; Ghanbarian-Alavijeh & Hunt, 2012; Hunt, 2001). We aim to perform soil classification using a large data set including 102 soil samples from the UNSODA database and assess the practical applications of the proposed theoretic method in soil classification and grouping.

2. Critical Path Analysis

Critical path analysis (CPA) is a promising approach from statistical physics (Ambegaokar et al., 1971; Pollak, 1972) that have been successfully used to model flow and transport in porous media, such as soils and rocks (Hunt & Sahimi, 2017; Hunt et al., 2014). Based on CPA, fluid flow in a network of pores of a wide range in size is controlled by larger pores whose magnitudes are greater than some critical pore size, which is defined as the smallest pore size required to let a fluid percolate through the network. In other words, pores with conductances greater than the critical value significantly contribute to flow, while low-conductance pores make trivial contribution to the overall transport, and, thus, they can be removed from flow analysis.

Let us map a soil into a network of pores of various sizes. To determine the critical pore conductance or size, one should first remove all pores and then replace them in their original locations in a decreasing order from the largest to the smallest size. As the first largest pores are replaced, there is still no percolating cluster. However, after a sufficiently large fraction of pores is replaced within the network, a sample-spanning cluster forms and the system starts percolating. That size resulting in the percolating cluster is the critical pore size.

CPA provides a theoretical framework to link hydraulic conductivity to critical pore size (Katz & Thompson, 1986). Under partially saturated conditions, as water saturation decreases the value of critical pore size and, consequently, hydraulic conductivity decrease as well. Hunt (2001) was first to apply concepts from CPA to model unsaturated hydraulic conductivity, $K(S_w)$. He estimated $K(S_w)$ from water retention curve and showed good agreement with experiments from the Hanford site, particularly at high water saturations. His approach was later generalized by Ghanbarian-Alavijeh and Hunt (2012) and Hunt et al. (2013) to be applicable to a broader range of porous media. A comprehensive review of CPA-based $K(S_w)$ models can be found in Ghanbarian et al. (2015) and Ghanbarian and Hunt (2017).

3. Theory of Soil Classification

Recently, Ghanbarian et al. (2019) developed a method to classify rocks based on their hydraulic properties under fully saturated conditions using saturated hydraulic conductivity and electrical conductivity data. In their approach, rocks with similar characteristic pore sizes are grouped into the same type using the hydraulic conductivity model of Johnson et al. (1986). In this study, we generalize their approach to unsaturated conditions by applying concepts from CPA. More specifically, we group soils that have similar critical pore radius, r_c, at the same effective water saturation S_e. To calculate r_c values, we use the unsaturated hydraulic conductivity measurements and convert them into critical pore radii using concepts from CPA (Ghanbarian, 2020; Hunt, 2001). For this purpose, one may utilize the following relationship (Ghanbarian & Hunt, 2017; Hunt, 2001)

$$\frac{K\left(S_w\right)}{K_s} = \left(\frac{r_c\left(S_w\right)}{r_c\left(S_w = 1\right)}\right)^{\gamma} \tag{1}$$

where $K(S_w)$ and K_s are respectively the unsaturated and saturated hydraulic conductivities, S_w is the water saturation, and $r_c(S_w)$ and $r_c(S_w = 1)$ are the critical pore radii under partially and fully saturated conditions, respectively. Ghanbarian, Hunt, and Daigle (2016) linked the exponent γ in Equation 1 to the surface fractal dimension, D_s, whose value can be determined by fitting the following pore-solid fractal (PSF) model to the soil water retention curve (Bird et al., 2000)

$$S_w = 1 - \frac{\beta}{\phi}\left(1 - \left(\frac{h}{h_a}\right)^{D_s - 3}\right), \quad h_a \le h \le h_{max} \tag{2}$$

in which h is the suction head, h_a is the air entry value, D_s is the surface fractal dimension, ϕ is the porosity, and $\phi \le \beta \le 1$ is the PSF model parameter.

Ghanbarian, Hunt, and Daigle (2016) theoretically showed that in isotropic media, one has

$$\gamma = 2\left(4 - D_s\right) - \frac{3 - D_s}{2D_s - 3} \tag{3}$$

In natural porous media, one has $2 \le D_s < 3$. Such a range leads to $2 < \gamma \lesssim 3$. In fact, as surface fractal dimension increases from 2 to 3, the value of γ decreases from ~3 to 2.

The value of effective water saturation can be determined from the following equation

$$S_e = \frac{S_w - S_{wc}}{1 - S_{wc}} \tag{4}$$

where S_{wc} is the critical water saturation at which the hydraulic conductivity vanishes. The value of S_{wc} may be approximated from the dry end of either soil water retention curve or unsaturated hydraulic conductivity.

4. Materials and Methods

4.1. Experiments

To assess the proposed soil classification approach, we used the data that Ghanbarian-Alavijeh and Hunt (2012) and Ghanbarian and Hunt (2017) analyzed in their studies. The selected data set from the UNSODA database (Nemes et al., 2001) included 102 soil samples from nine soil texture classes. For the distribution of these samples within the soil texture triangle; see Figure 1 in Ghanbarian-Alavijeh and Hunt (2012). For each soil sample both the water retention and unsaturated hydraulic conductivity curves were available shown in Figure 1. Most of the water retention curves contained more than 10 measured data points from full saturation to 15,000 cm H_2O. The PSF water retention model, Equation 2, was fitted by Ghanbarian-Alavijeh and Hunt (2012) who reported the optimized values of D_s, β, and h_a for each soil sample (see Table 1). We calculated the value of the exponent γ from D_s via Equation 3. Accurate determination of S_{wc} requires precise measurements of water retention or unsaturated hydraulic conductivity curve at the dry end (or high suction heads). However, measuring the unsaturated hydraulic conductivity curve near the critical water saturation is time-consuming, particularly in fine-textured soils. Because of very low hydraulic conductivity, experiment time might be not adequately long enough to reach equilibrium conditions. Following Ghanbarian-Alavijeh and Hunt (2012), the critical water saturation, S_{wc}, was *roughly*

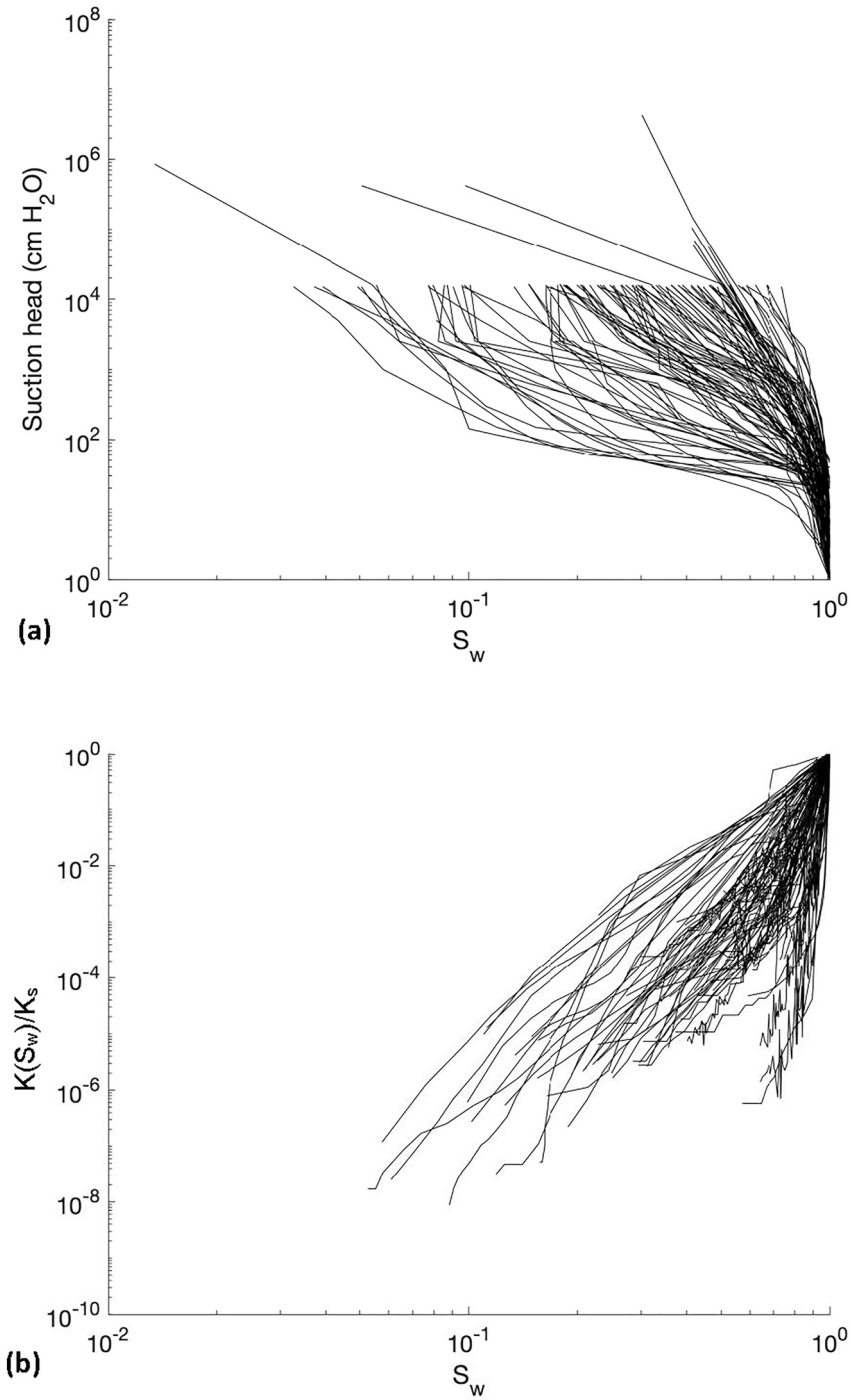

Figure 1. (a) The water retention and (b) the unsaturated hydraulic conductivity curves for the 102 soil samples from the UNSODA database studied here.

approximated from the water saturation at the maximum suction head (in most cases at 15,000 cm H_2O). We consequently calculated the effective water saturation, S_e, from Equation 4.

The critical pore radii at different water saturations were determined from the measured unsaturated hydraulic conductivity values using Equation 1. To compare the values of $r_c(S_w)$ and $r_c(S_w = 1)$ from different soils at the same S_e value, the $K(S_w)/K_s - S_w$ curves were accordingly converted into the $r_c(S_w)/r_c(S_w = 1) - S_e$ ones.

Table 1
The Average Values of the Pore-Solid Fractal Model Parameters for Each Soil Texture Class

Soil texture	No. of samples	h_a (cm)	β	D_s	γ	R^2
Sand	24	18.7	0.45	2.649	2.55	0.97
Loamy sand	5	20.6	0.44	2.659	2.53	0.95
sandy loam	18	21.0	0.73	2.906	2.15	0.97
Loam	10	18.0	0.82	2.939	2.10	0.97
Silt loam	41	27.3	0.93	2.932	2.11	0.95
Sandy clay loam	7	10.7	0.90	2.963	2.06	0.99
Silty clay loam	3	52.1	0.96	2.962	2.06	0.98
Silty clay	4	55.2	0.95	2.949	2.08	0.98
Clay	3	28.8	0.96	2.965	2.06	0.94

Note. The optimized parameters h_a, β, and D_s were reported by Ghanbarian-Alavijeh and Hunt (2012). The value of γ was determined from the D_s via Equation 3.

4.2. Curve Clustering

Various clustering approaches have been developed in the literature, such as *k*-means, single linkage, and simulated annealing (Maulik & Bandyopadhyay, 2002). Most, however, classifies data based on individual measurements represented by single values. For example, the *k*-means clustering method partitions observations into *k* clusters in each of which samples belong to the same cluster with the nearest mean. However, it is known that the *k*-means approach has the limitation of getting stuck at suboptimal configurations, depending on the choice of the initial cluster centers. However, the simulated annealing method can overcome this limitation because it is capable of handling local optima (Maulik & Bandyopadhyay, 2002).

Curve clustering is a probabilistic framework that allows for the joint clustering and continuous alignment of sets of curves in curve space (Gaffney, 2004). It integrates probabilistic alignment models with model-based curve clustering algorithms. The latter allows for the derivation of consistent learning algorithms for the joint clustering-alignment problem. More specifically, the curve clustering approach uses two regression mixing models: (a) polynomial regression (lrm) and (b) spline regression (srm) to insert similar curves in the same cluster. It is based on a regression mixture equation with up to four transformations, given by

$$y = c\left[ax + b\right]B + d + e \tag{5}$$

where the values within the square brackets represent the transformed regression matrix, c is related to the scaling in the measurement space, d is linked to translation in the measurement space, a is related to the scaling in time, and b is related to translation in time. In the present study, we assumed that S_e played the role of time, while $r_c(S_w)/r_c(S_w = 1)$ was the measurement space, or S_e and $r_c(S_w)/r_c(S_w = 1)$ were represented by the horizontal and vertical axes of a plot. Using these transformation parameters, we attempted to cluster all the data into representative soil classes as discussed earlier. We applied the open-source toolbox developed by Gaffney (2004). It is a MATLAB toolbox for clustering curve data using various probabilistic curve-based mixture models. We attempted different regression models (e.g., lrm and srm), transformation models (i.e., Equation 5 with different parameters), polynomial orders (between 1 and 4), and the number of clusters within the curve clustering method in this study.

5. Results and Discussion

Figure 1b shows the unsaturated hydraulic conductivity curves, on the log-log scale, measured on 102 soil samples from the UNSODA database. As can be seen, the hydraulic conductivity spans nearly eight orders of magnitude, while the water saturation mainly ranges between 0.1 and 1. This indicates the soil samples studied here cover a wide range of the unsaturated hydraulic conductivity measurements. On one hand, there are fine-textured soils for which the hydraulic conductivity varies mildly with the water saturation. On the other hand, there exist coarse-textured samples with sharp decrease in $K(S_w)$ when S_w drops from 1 to 0.7 probably due to the presence of macropores (Jarvis, 2008). From experimental viewpoint, reaching equilibrium at very low water saturations near S_{wc} where $K(S_w)$ tends to vanish is time-consuming. Figure 1b clearly shows that most hydraulic conductivity measurements are far above the critical water saturation justifying S_{wc} estimation from the water retention curves.

Figure 2 displays the histograms of D_s and γ. The average values of D_s and γ reported in Table 1 for each soil texture indicate a relatively wide range. We found the smallest and greatest average $D_s = 2.649$ and 2.965, respectively in the sand and clay textural classes. The obtained results show a clear trend between D_s and soil texture; D_s value increases with an increase of clay content. This is in accord with earlier results reported by Huang and Zhang (2005), Wang et al. (2005), Ghanbarian-Alavijeh and Millán (2009), and many others. Since γ is linked to D_s via Equation 3, one should expect γ to be greater in coarse-textured soils than that in

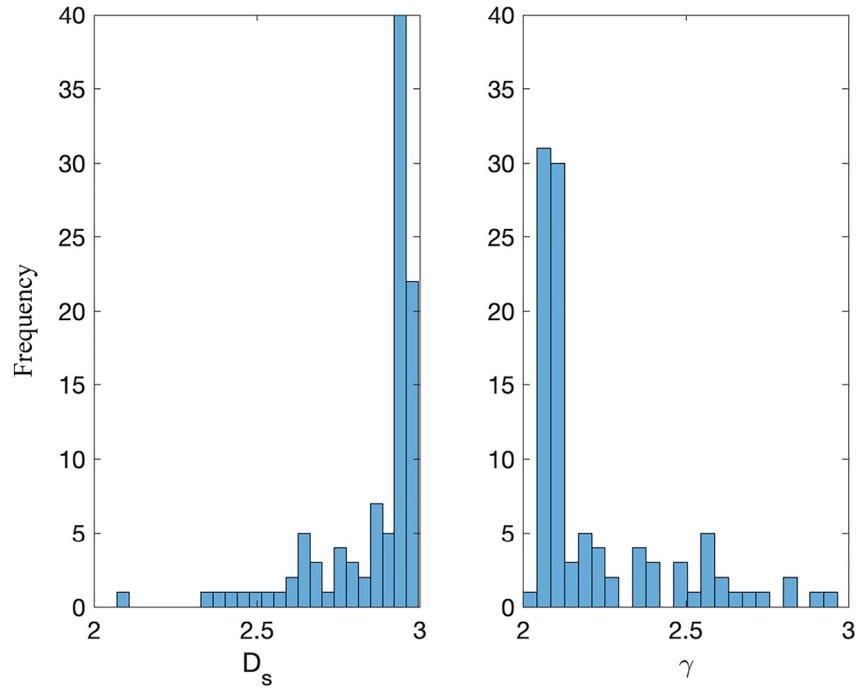

Figure 2. The histograms of the D_s and γ values for the 102 soil samples from the UNSODA database used in this study.

fine-textured ones (see Table 1). We found the average $\gamma = 2.55$ for the sand and $\gamma = 2.06$ for the clay textural classes.

We show the normalized critical pore radius, $r_c(S_w)/r_c(S_w = 1)$, against the effective water saturation, S_e, on the logarithmic scale in Figure 3. The $r_c(S_w)/r_c(S_w = 1)$ spans near three orders of magnitude. This is because $K(S_w)$ is linked to $r_c(S_w)$ through the power law given in Equation 1. The exponent γ in Equation 1 controls variations in $K(S_w)$. For example, if $r_c(S_w)/r_c(S_w = 1)$ spans two orders of magnitude, one should expect $K(S_w)/K_s$ to span four orders of magnitude, if $\gamma = 2$.

For curve clustering, we attempted both the polynomial regression (lrm) and the spline regression (srm) models with each available transformation parameters along with different polynomial orders (between 1 and 4) and the number of clusters. We found the most reliable and consistent results were obtained using a polynomial regression with the $y = \left[x \right] B + d$ transformation as well as an order of three and eight clusters. Other polynomial orders produced inconsistent results for example, grouping all soil samples into a singular group regardless of the number of clusters specified. Cluster numbers smaller than eight grouped unlike curves into similar classes, while numbers greater than eight resulted in grouping similar curves into different groups.

Results of the curve clustering method and soil classification are presented in Figure 4. As can be seen, soils with similar $r_c(S_w)/r_c(S_w = 1)-S_e$ curves were grouped into the same class. The ratio $r_c(S_w)/r_c(S_w = 1)$, indicating the normalized critical pore radius at different effective water saturations, represents some hydraulic characteristic that controls fluid flow based on the CPA. We found eight main soil classes using the proposed approach in combination with the curve clustering method (Figure 4). This means in the entire database analyzed here, there exists eight main classes that are substantially different in terms of pore space and contact angle properties.

Interestingly, our results and the number of soil classes are not greatly different from those of Twarakavi et al. (2010) who investigated the numbers of soil hydraulic classes. They used the ROSETTA software package to determine the optimal soil hydraulic classifications and their associated uncertainties. Twarakavi et al. (2010) varied the number of soil hydraulic classes from 2 to 30 and concluded that the optimal number of soil hydraulic classes was twelve.

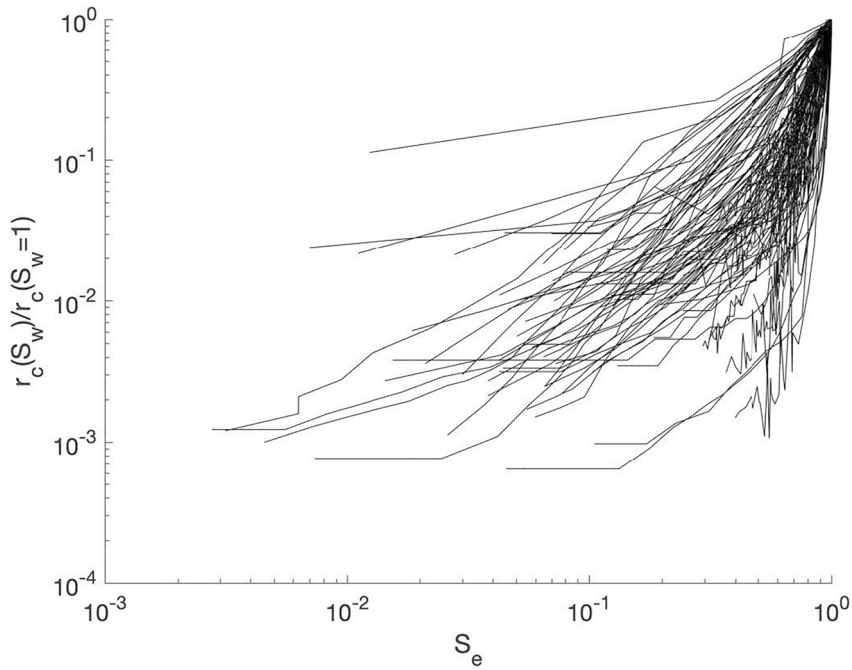

Figure 3. The $r_c\left(S_w\right)/r_c\left(S_w = 1\right) - S_e$ curves for the 102 soil samples from the UNSODA database studied here.

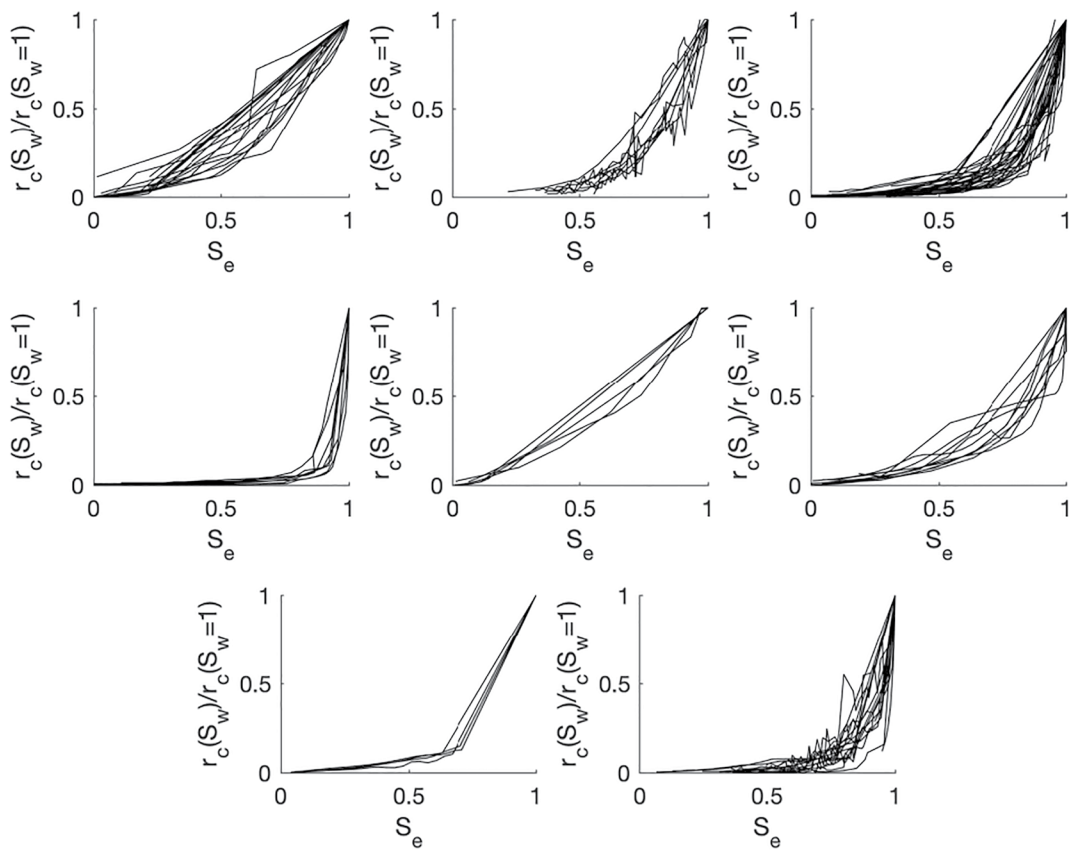

Figure 4. Eight different soil groups detected by the curve clustering method and using the $r_c\left(S_w\right)/r_c\left(S_w = 1\right) - S_e$ curves.

Table 2
Soil Samples With Their Corresponding UNSODA Soil Codes Detected in Each Soil Class Using the Proposed Approach

Soil class	UNSODA soil code
1	1,463, 1,464, 4,000, 4,001, 4,020, 4,021, 4,050, 4,060, 4,061, 4,161, 4,523, 4,590, 4,610, 4,612, 4,651, 4,660
2	1,280, 1,281, 1,380, 1,391, 1,465, 2,530, 2,550, 4,602, 4,870
3	1,392, 1,467, 2,551, 2,552, 3,120, 3,361, 3,370, 3,371, 3,380, 3,390, 3,392, 3,393, 4,011, 4,031, 4,041, 4,052, 4,070, 4,081, 4,092, 4,100, 4,111, 4,120, 4,121, 4,140, 4,151, 4,170, 4,171, 4,181, 4,670, 4,671, 4,672, 4,673
4	4,032, 4,033, 4,101, 4,172, 4,182, 4,183, 4,680, 4,681
5	1,461, 1,462, 1,466, 4,592, 4,661
6	3,395, 4,030, 4,080, 4,110, 4,160, 4,450, 4,620, 4,621, 4,622, 4,650
7	4,010, 4,130, 4,141, 4,162
8	1,360, 1,361, 1,362, 1,370, 1,381, 1,383, 1,390, 1,400, 1,490, 2,531, 2,532, 2,540, 2,541, 3,360, 4,062, 4,071, 4,510, 4,530

	Class							
USDA soil texture	1	2	3	4	5	6	7	8
Sand	9	0	4	0	4	1	1	0
Sandy Loam	1	2	8	1	0	2	2	3
Loamy Sand	1	1	0	0	0	0	0	1
Loam	3	3	1	1	1	3	0	4
Silt Loam	2	3	15	4	0	4	1	5
Silt	0	0	1	0	0	0	0	0
Sandy Clay Loam	0	0	0	0	0	0	0	0
Clay Loam	0	0	0	0	0	0	0	1
Silty Clay Loam	0	0	0	0	0	0	0	0
Sandy Clay	0	0	0	0	0	0	0	0
Silty Clay	0	0	1	0	0	0	0	2
Clay	0	0	2	2	0	0	0	2

Soils included in each detected class, denoted by their UNSODA database code, and their textural classes based on the USDA classification are listed in Table 2. For classes 5 and 7, we only found five and four samples, respectively, while for class 3, there exist 32 samples. We compared our soil classification results with those from Sadeghi et al. (2016) who classified 26 soils from the UNSODA database using a joint scaling factor based on air entry value, $h_{a,}$ and saturated hydraulic conductivity, K_s. As we show in Table 2, we found that soil samples 3,120, 4,030, and 4,870 belonged to classes 3, 6, and 2, respectively, whereas based on Sadeghi et al. (2016) analyses those samples grouped into the same class (their class 3). Similarly, soils with codes 3,360, 3,380, and 3,395 belonged respectively to classes 8, 3, and 6. However, Sadeghi et al. (2016) grouped them in their class 6. The main reason for discrepancies between our results and those reported by Sadeghi et al. (2016) is that those authors used one data point from the water retention curve that is, h_a and another data point from the unsaturated hydraulic conductivity curve that is, K_s to define their joint factor and scale water retention and unsaturated hydraulic conductivity curves. However, we use the entire water retention and unsaturated hydraulic conductivity curves to classify soils.

In the literature, grouping has been widely used as a strategy to develop more accurate pedotransfer functions by identifying soils with similar characteristics (Bruand, 2004). As stated earlier, soil textural properties along with bulk density have been widely applied to group soils. However, such classifications do not take into account the effect of solid-fluid characteristics, such as contact angle, that might influence hydraulic properties of soils. It has been well documented that the unsaturated hydraulic conductivity depends on wettability and contact angle (Diamantopoulos & Durner, 2013; Sahimi, 2011). Because the contact angle affects the unsaturated hydraulic conductivity measurements, the soil classification method proposed in

this study indirectly incorporates the effect of wettability on grouping soils through converting the $K(S_w)/K_s - S_w$ curves into the $r_c(S_w)/r_c(S_w = 1) - S_e$ ones.

Under hydrophilic conditions, water as the wetting and air as the non-wetting fluids occupy the smallest and largest pores, respectively. The smallest pores are, however, ignored within the CPA framework, whereas the largest pores are taken into account. Consequently, our proposed soil classification approach cannot be used to classify soils using air permeability curves under hydrophilic conditions or using unsaturated hydraulic conductivity curves under hydrophobic circumstances. Alternatively, one may apply another upscaling technique from statistical physics, such as the effective-medium approximation (Ghanbarian, Sahimi, & Daigle, 2016; Levine & Cuthiell, 1986) for soil classification purposes.

In the past several years, machine learning has become a widely used tool to analyze data in science and engineering. Grouping soils has practical applications to data classification- and regression-type problems in machine learning (Aggarwal, 2014; Matloff, 2017), particularly when a large number of data is available. Unlike the existing soil classification methods, the proposed approach is on the basis of soil hydraulic conductivity and its saturation-dependent curve, which indirectly takes the effect of contact angle into account. This makes our approach an appropriate classification method for the unsaturated zone, which is critical in the partitioning of precipitation to runoff, evaporation, storage, or recharge.

6. Conclusion

Grouping soils has broad practical applications in hydrology and soil science. Accordingly, we proposed a new theoretic approach using concepts from critical path analysis to classify soils based on similarities in the critical pore radius at the same effective water saturation. To evaluate our approach, we selected 102 soil samples for which both the water retention and unsaturated hydraulic conductivity curves were available. We first converted the measured $K(S_w)/K_s - S_w$ curves into the $r_c(S_w)/r_c(S_w = 1) - S_e$ ones, then applied the curve clustering method to group soils, and found eight soil classes. Results also demonstrated discrepancies among different soil classification methods, and the importance of hydraulic flow characteristics when classifying soils. Our number of soil classes align closely with previously determined soil hydraulic classes. This indicates that CPA can be an effective tool in classifying soil classes when the water retention and unsaturated hydraulic conductivity curves are available. Further investigations are required to expand the developed soil classification approach for air permeability measurements.

Data Availability Statement

The data used in this studied are available at: http://www.hydroshare.org/resource/53be7b9e275746c0868956bc46781204.

Acknowledgment
B. Ghanbarian is grateful to Kansas State University for supports through faculty start-up funds.

References

Aggarwal, C. C. (2014). *Data classification: Algorithms and applications*. CRC Press.

Ambegaokar, V., Halperin, B. I., & Langer, J. S. (1971). Hopping conductivity in disordered systems. *Physical Review B: Condensed Matter*, *4*(8), 2612–2620. https://doi.org/10.1103/physrevb.4.2612

Bird, N. R. A., Perrier, E., & Rieu, M. (2000). The water retention function for a model of soil structure with pore and solid fractal distributions. *European Journal of Soil Science*, *51*(1), 55–63. https://doi.org/10.1046/j.1365-2389.2000.00278.x

Bruand, A. (2004). Preliminary grouping of soils. *Developments in Soil Science*, *30*, 159–174. https://doi.org/10.1016/s0166-2481(04)30010-3

Dennis, E. J., & Berbery, E. H. (2021). The role of soil texture in local land surface–atmosphere coupling and regional climate. *Journal of Hydrometeorology*, *22*, 313–330. https://doi.org/10.1175/jhm-d-20-0047.1

Devine, S. M., Steenwerth, K. L., & O'Geen, A. T. (2021). A regional soil classification framework to improve soil health diagnosis and management. *Soil Science Society of America Journal*, *85*, 361–378. https://doi.org/10.1002/saj2.20200

Diamantopoulos, E., & Durner, W. (2013). Physically-based model of soil hydraulic properties accounting for variable contact angle and its effect on hysteresis. *Advances in Water Resources*, *59*, 169–180. https://doi.org/10.1016/j.advwatres.2013.06.005

Gaffney, S. (2004). *Probabilistic curve-aligned clustering and prediction with regression mixture models* (Doctoral dissertation). University of California.

García-Gutiérrez, C., Pachepsky, Y., & Martín, M. Á. (2018). Saturated hydraulic conductivity and textural heterogeneity of soils. *Hydrology and Earth System Sciences*, *22*, 3923–3932. https://doi.org/10.5194/hess-22-3923-2018

Ghanbarian, B. (2020). Applications of critical path analysis to uniform grain packings with narrow conductance distributions: II. Water relative permeability. *Advances in Water Resources*, *137*, 103524. https://doi.org/10.1016/j.advwatres.2020.103524

Ghanbarian, B., & Hunt, A. G. (2017). Improving unsaturated hydraulic conductivity estimation in soils via percolation theory. *Geoderma*, *303*, 9–18. https://doi.org/10.1016/j.geoderma.2017.05.004
Ghanbarian, B., Hunt, A. G., & Daigle, H. (2016). Fluid flow in porous media with rough pore-solid interface. *Water Resources Research*, *52*(3), 2045–2058. https://doi.org/10.1002/2015WR017857
Ghanbarian, B., Hunt, A. G., Skinner, T. E., & Ewing, R. P. (2015). Saturation dependence of transport in porous media predicted by percolation and effective medium theories. *Fractals*, *23*(1), 1540004. https://doi.org/10.1142/S0218348X15400046
Ghanbarian, B., Lake, L. W., & Sahimi, M. (2019). Insights into rock typing: A critical study. *SPE Journal*, *24*, 230–242. https://doi.org/10.2118/191366-pa
Ghanbarian, B., Sahimi, M., & Daigle, H. (2016). Modeling relative permeability of water in soil: Application of effective-medium approximation and percolation theory. *Water Resources Research*, *52*(7), 5025–5040. https://doi.org/10.1002/2015WR017903
Ghanbarian-Alavijeh, B., & Hunt, A. G. (2012). Unsaturated hydraulic conductivity in porous media: Percolation theory. *Geoderma*, *187–188*, 77–84. https://doi.org/10.1016/j.geoderma.2012.04.007
Ghanbarian-Alavijeh, B., & Millán, H. (2009). The relationship between surface fractal dimension and soil water content at permanent wilting point. *Geoderma*, *151*(3–4), 224–232. https://doi.org/10.1016/j.geoderma.2009.04.014
Huang, G., & Zhang, R. (2005). Evaluation of soil water retention curve with the pore-solid fractal model. *Geoderma*, *127*(1–2), 52–61. https://doi.org/10.1016/j.geoderma.2004.11.016
Hunt, A., Ewing, R., & Ghanbarian, B. (2014). *Percolation theory for flow in porous media*. Springer.
Hunt, A. G. (2001). Applications of percolation theory to porous media with distributed local conductances. *Advances in Water Resources*, *24*(3–4), 279–307. https://doi.org/10.1016/S0309-1708(00)00058-0
Hunt, A. G., Ghanbarian, B., & Saville, K. C. (2013). Unsaturated hydraulic conductivity modeling for porous media with two fractal regimes. *Geoderma*, *207–208*(1), 268–278. https://doi.org/10.1016/j.geoderma.2013.05.023
Hunt, A. G., & Sahimi, M. (2017). Flow, transport, and reaction in porous media: Percolation scaling, critical-path analysis, and effective medium approximation. *Reviews of Geophysics*, *55*(4), 993–1078. https://doi.org/10.1002/2017RG000558
Jarvis, N. (2008). Near-saturated hydraulic properties of macroporous soils. *Vadose Zone Journal*, *7*(4), 1256–1310. https://doi.org/10.2136/Vzj2008.0065
Johnson, D. L., Koplik, J., & Schwartz, L. M. (1986). New pore-size parameter characterizing transport in porous media. *Physical Review Letters*, *57*(20), 2564–2567. https://doi.org/10.1103/physrevlett.57.2564
Katz, A. J., & Thompson, A. H. (1986). Quantitative prediction of permeability in porous rock. *Physical Review B: Condensed Matter*, *34*(11), 8179–8181. https://doi.org/10.1103/physrevb.34.8179
Klute, A., & Wilkinson, G. E. (1958). Some tests of the similar media concept of capillary flow: I. Reduced capillary conductivity and moisture characteristic data. *Soil Science Society of America Proceedings*, *22*, 278–281. https://doi.org/10.2136/sssaj1958.03615995002200040002x
Levine, S., & Cuthiell, D. L. (1986). Relative permeabilities in two-phase flow through porous media: An application of effective medium theory. *Journal of Canadian Petroleum Technology*, *25*(5), 74–84. https://doi.org/10.2118/86-05-10
Matloff, N. (2017). *Statistical regression and classification: From linear models to machine learning*. CRC Press.
Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1650–1654. https://doi.org/10.1109/tpami.2002.1114856
Miller, E. E., & Miller, R. D. (1956). Physical theory for capillary flow phenomena. *Journal of Applied Physics*, *27*, 324–332. https://doi.org/10.1063/1.1722370
Nemes, A., Schaap, M. G., Leij, F. J., & Wösten, J. H. M. (2001). Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology*, *251*(3–4), 151–162. https://doi.org/10.1016/S0022-1694(01)00465-6
Nguyen, P. M., De Pue, J., Van Le, K., & Cornelis, W. (2015). Impact of regression methods on improved effects of soil structure on soil water retention estimates. *Journal of Hydrology*, *525*, 598–606. https://doi.org/10.1016/j.jhydrol.2015.04.014
Nielsen, D. R., Hopmans, J. W., & Reichardt, K. (1998). An emerging technology for scaling field soil-water behaviour. In G. Sposito (Ed.), *Scale dependence and scale invariance in hydrology*, (pp. 136–166). Cambridge University Press. https://doi.org/10.1017/cbo9780511551864.006
NRCS. (1986). *Urban hydrology for small watersheds (No. 55)*. Engineering Division, Soil Conservation Service, US Department of Agriculture.
Pachepsky, Y., & Park, Y. (2015). Saturated hydraulic conductivity of US soils grouped according to textural class and bulk density. *Soil Science Society of America Journal*, *79*(4), 1094–1100. https://doi.org/10.2136/sssaj2015.02.0067
Pachepsky, Y. A., & Rawls, W. J. (1999). Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Science Society of America Journal*, *63*(6), 1748–1757. https://doi.org/10.2136/sssaj1999.6361748x
Pachepsky, Y. A., Rawls, W. J., & Lin, H. S. (2006). Hydropedology and pedotransfer functions. *Geoderma*, *131*(3–4), 308–316. https://doi.org/10.1016/j.geoderma.2005.03.012
Pollak, M. (1972). A percolation treatment of dc hopping conduction. *Journal of Non-Crystalline Solids*, *11*(1), 1–24. https://doi.org/10.1016/0022-3093(72)90304-3
Rawls, W. J., & Pachepsky, Y. A. (2002a). Soil consistence and structure as predictors of water retention. *Soil Science Society of America Journal*, *66*, 1115–1126. https://doi.org/10.2136/sssaj2002.1115
Rawls, W. J., & Pachepsky, Y. A. (2002b). Using field topographic descriptors to estimate soil water retention. *Soil Science*, *167*, 423–435. https://doi.org/10.1097/00010694-200207000-00001
Sadeghi, M., Ghahraman, B., Warrick, A. W., Tuller, M., & Jones, S. B. (2016). A critical evaluation of the Miller and Miller similar media theory for application to natural soils. *Water Resources Research*, *52*, 3829–3846. https://doi.org/10.1002/2015wr017929
Sahimi, M. (2011). *Flow and transport in porous media and fractured rock: From classical methods to modern approaches* (2nd ed.). Wiley-VCH.
Schroth, M. H., Ahearn, S. J., Selker, J. S., & Istok, J. D. (1996). Characterization of Miller-similar silica sands for laboratory hydraulic studies. *Soil Science Society of America Journal*, *60*, 1331–1339. https://doi.org/10.2136/sssaj1996.03615995006000050007x
Twarakavi, N. K. C., Šimůnek, J., & Schaap, M. G. (2010). Can texture-based classification optimally classify soils with respect to soil hydraulics? *Water Resources Research*, *46*, W01501. https://doi.org/10.1029/2009WR007939
Vereecken, H., Kasteel, R., Vanderborght, J., & Harter, T. (2007). Upscaling hydraulic properties and soil water flow processes in heterogeneous soils. *Vadose Zone Journal*, *6*(1), 1–28. https://doi.org/10.2136/vzj2006.0055
Wang, K., Zhang, R., & Wang, F. (2005). Testing the pore-solid fractal model for the soil water retention function. *Soil Science Society of America Journal*, *69*(3), 776–782. https://doi.org/10.2136/sssaj2004.0247