Saturated hydraulic conductivity prediction from microscopic pore geometry measurements and neural network analysis

I. Lebron, M. G. Schaap, and D. L. Suarez

U.S. Salinity Laboratory, ARS, USDA, Riverside, California

Abstract. Traditional models to describe hydraulic properties in soils are constrained by the assumption of cylindrical capillarity to account for the geometry of the pore space. This study was conducted to develop a new methodology to directly measure the porosity and its microscopic characteristics. The methodology is based on the analysis of binary images collected with a backscattered electron detector from thin sections of soils. Pore surface area, perimeter, roughness, circularity, and maximum and average diameter were quantified in 36 thin sections prepared from undisturbed soils. Saturated hydraulic conductivity K_{sat}, particle size distribution, particle density, bulk density, and chemical properties were determined on the same cores. We used the Kozeny-Carman equation and neural network and bootstrap analysis to predict a formation factor from microscopic, macroscopic, and chemical data. The predicted K_{sat} was in excellent agreement with the measured K_{sat} ($R^2 = 0.91$) when a hydraulic radius r_H defined as pore area divided by pore perimeter and the formation factor were included in the Kozeny-Carman equation.

1. Introduction

Flow and transport of water and solutes in soils are controlled by size, geometry, and characteristics of the soil porosity. Most of the characteristics of soil pores are microscopic, such as roughness and circularity. Conventional models of liquid distribution, flow, and solute transport rely solely on cylindrical capillarity, ignoring the role of surface area, angularity, and connectivity. It is known that the geometry of the soil pores is very irregular; for example, Figure 1 shows a thin section from an undisturbed sample of Gilman silt loam soil. We identify polygonal and angular pores associated with feldspars and quartz while the more elongated pores are, in general, associated with clays. These different shapes are important because the water retention in the pores depends on the angularity and roughness of the pores [*Tuller et al.*, 1999]. Also, elongated pores may be more susceptible to the effect of the chemical composition of the soil water since they are associated with clays, which have colloidal properties.

If we look closely at one of the features in Figure 1, we see a montmorillonite domain (Figure 2). The arrangement of this organized structure modifies with changes in the chemical composition, as has been demonstrated in previous studies with electrophoretic mobility for montmorillonites [*Shainberg and Otoh*, 1968] and illites [*Lebron et al.*, 1993]. The hypothesis of this study is that the self-arrangement of the particles in the soil is not totally random and if we understand the forces determining this self-arrangement, we will be in a better position to model pore space and, consequently, water flow.

The use of thin sections to measure grain sizes has been performed for more than 60 years in petrological studies [*Friedman*, 1958, 1962]. The advance of new technologies allows the use of more powerful microscopes and sophisticated software packages to quickly process thousands of measurements. Scanning electron microscopy and image analysis have

been successfully applied to the measurement of porosity in rocks by *Chretien and Bisdom* [1983], *Schoonderbeek et al.* [1983], *Berryman* [1998], *Berryman and Blair* [1987], *Ehrlich et al.* [1991], and *Blair et al.* [1996] and in soils by *Bruand et al.* [1996] and *Vogel* [1997].

Image-based technology provides unique information about pore geometry by direct measurement. While other techniques are used to obtain indirect or deducted values using a number of assumptions, microscopy and image analysis supply actual values of the pore space. The size of the pores resolved in the image depends on the magnification and the size of the pixels. *Berryman and Blair* [1987], using correlation functions, developed a methodology to derive the characteristic length at which the physical properties of the heterogeneous soil microstructure are spatially correlated. More recently, *Blair et al.* [1996], with a two-point correlation function, determined the adequate image resolution to obtain image parameters consistent with those used in a simple flow model, such as the Kozeny-Carman equation, for prediction of the permeability.

Neural network techniques have been used to predict water retention properties in soils by *Pachepsky et al.* [1996] and *Schaap and Bouten* [1996] using macroscopic parameters. We can find no examples in the literature of the use of neural networks in combination with pore microfeatures to predict flow in soils.

The objectives of this study are as follows: (1) to establish a methodology using a two-point correlation function and image analysis software to analyze soil pore space and pore geometry from microphotographs and (2) to interpret these data in the context of the model proposed by Kozeny and Carman using neural network and bootstrap analysis to predict soil hydraulic properties.

2. Materials and Methods

The soil used in this study was Gilman silt loam soil, from Coachella Valley, California. The field, 9.7 ha of saline sodic soil, was monitored and mapped in a previous survey for electrical conductivity (EC) and sodium adsorption ratio (SAR)
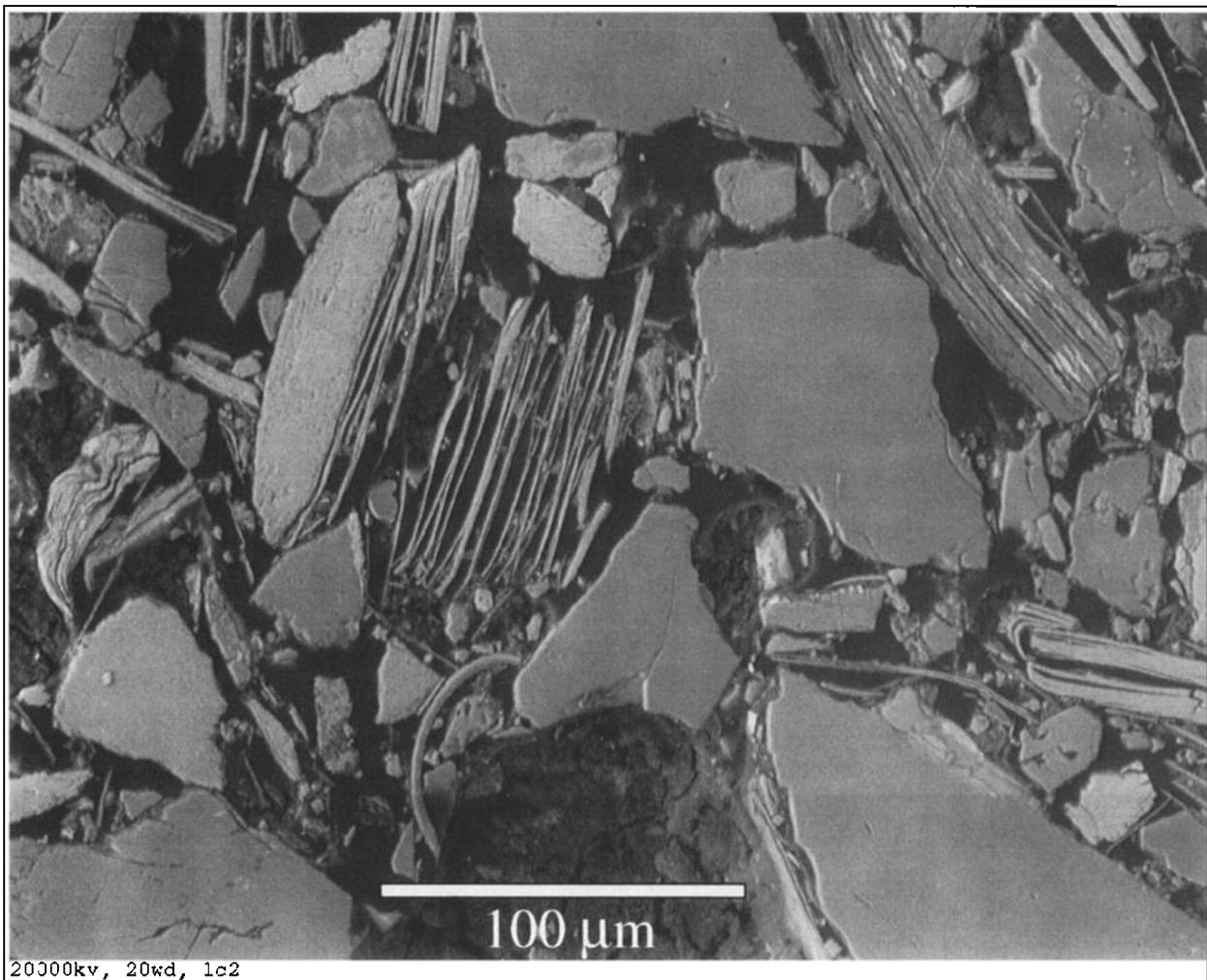
Figure 1. Scanning electron micrograph from a thin section of an undisturbed sample from Gilman silt loam soil, Coachella Valley, California.

(SAR $= Na^+/(Ca^{2+} + Mg^{2+})^{0.5}$, where the cation concentration is expressed in mmol L^{-1}). A sampling scheme was established to cover a maximum range of EC and SAR which resulted in 36 undisturbed soil samples. The undisturbed soil cores were 12 cm in diameter by 12 cm in height and were collected from the top 25 cm of the soil by hand, to avoid excessive mechanical compression of the natural structure. Saturated hydraulic conductivity K_{sat} was measured on all soil cores using Colorado River water as eluent (water used for irrigation on those fields); the chemical composition of the Colorado River water is as follows: EC, 1.15 dS m^{-1}; pH, 8.15; SAR, 2.8; HCO_3^-, 2.8; Cl^-, 3.1; SO_4^{2-}, 5.8; Na^+, 5.0; K^+, 0.12; Ca^{2+}, 4.0; and Mg^{2+}, 2.5; where ion concentrations are in $mmol_c$ L^{-1}.

Samples were saturated by first wetting by capillary rise from below, then gradually raising the water level until water ponded on the surface; K_{sat} was measured with the constant head method. After the K_{sat} was measured, bulk density ρ_B was determined, and the air-dried cores were cut vertically with a knife (in the direction of the water flow) into two equal parts. One part was used to prepare thin sections, and the other half was used to analyze for particle density ρ_d [*Blake and Hartge*, 1986], particle size distribution [*Gee and Bauder*, 1986], and

chemical composition (using inductively coupled plasma emission spectroscopy). Because of the scarcity of sample, EC, pH, and SAR were measured in a combination of the three nearest samples; the same value appears for the three combined soils. The results of the analyses are shown in Table 1. Thin sections were prepared by impregnation of the samples with epoxy EPO-TEK 301 (Epoxy Technology Inc., Billerica, Massachusetts). After hardening, a thin section 3.5 × 2.5 cm was cut at the plane perpendicular to the water flow, mounted on a glass slide, and polished. The polishing process was done with a series of diamond polishers to avoid the introduction of contaminants and in the absence of water to preserve soluble minerals.

Thin sections were observed in a scanning electron microscope (SEM) (AMRAY 3200, AMRAY Inc., Bedford, Massachusetts) with the backscatter electron detector. The intensity of the backscattered electrons (BE) is a function of the atomic weight of the element, with heavier elements having higher backscattering properties. The result is that elements with higher atomic weight give images that are brighter than lighter elements in the sample. Charging effects are counteracted by using low vacuum in the specimen chamber and by the introduction of air molecules to dissipate the charge accumulated
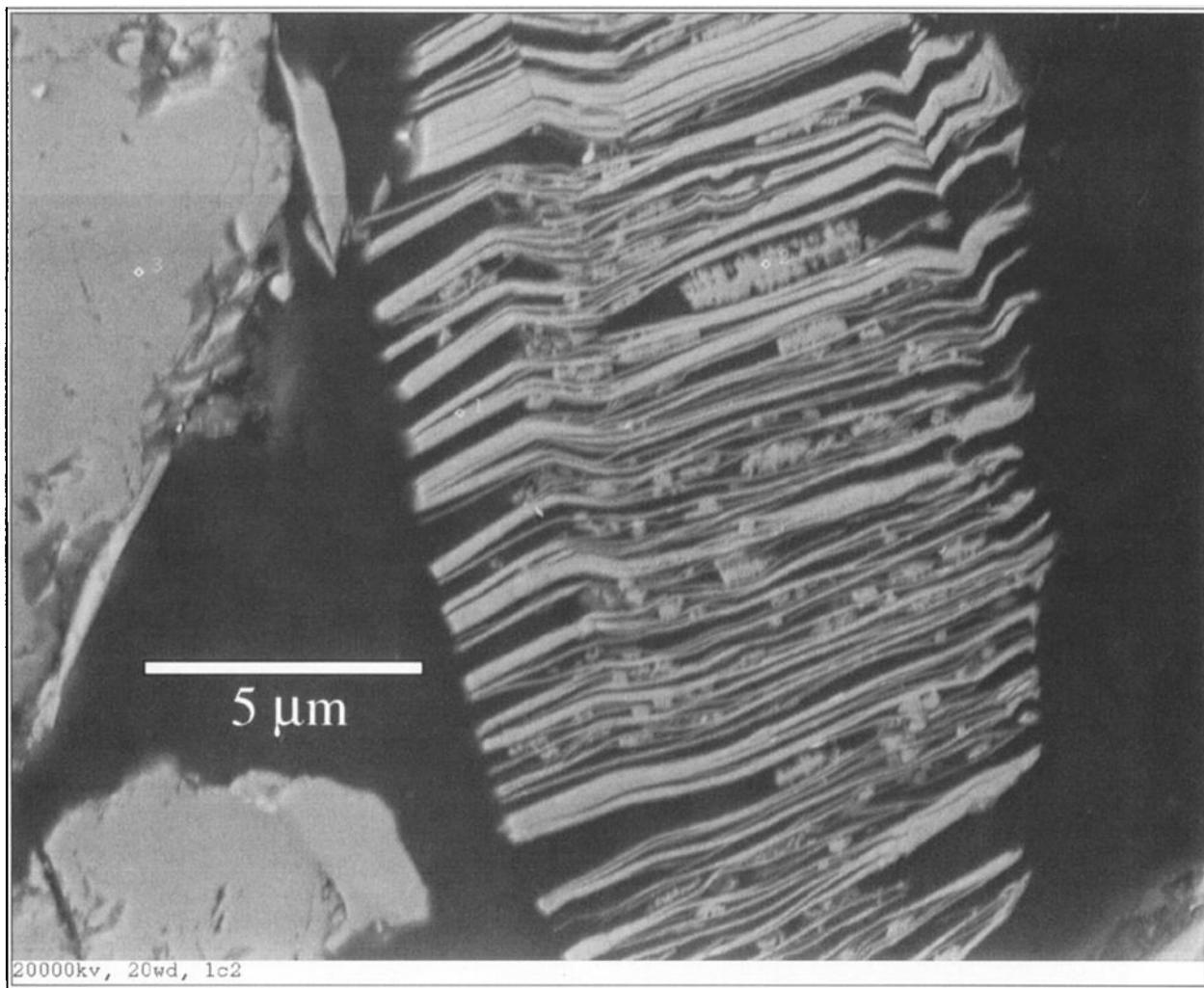
Figure 2. Scanning electron micrograph of a montmorillonite domain.

on the surface of the sample. No coating of the sample is needed. The selection of the magnification is made with a two-point correlation function (see section 3.2) and the quantification and classification of pore spaces with commercially available software (Princeton Gamma-Tech Inc., Princeton, New Jersey). The image collected by the scanning electron microscope in the backscatter electron detector (1024×800 pixels) is transformed into a binary image, and maximum, minimum, and average diameter (D_{max}, D_{min}, and D_{avg}, respectively), surface area A, and pore perimeter P are quantified by directly measuring the number of pixels that conform each feature in the binary image. Roughness R is calculated by $R = P/(\pi D_{avg})$, and the circularity C is calculated by $C = \pi D_{max}^2/(4A)$.

To increase the size of the sampling area, we collected 10 pictures from each thin section distributed in a regular grid; the total area sampled is equivalent to 2.5 mm^2. Each of the pictures was processed; average parameters are presented in Table 1.

3. Development of the Microscopic Method

3.1. Determination of the Threshold Grey Level

The images collected with the BE detector contain a grey scale that can be represented by a histogram. If the brightness and contrast of the image are adequate, the histogram presents a bimodal distribution, since the carbon-based epoxy displays as a black color and the soil minerals are bright in a wide spectra of grey colors. In general, the quality of the histogram for soil thin sections is dependent on the clay content; the higher the clay content, the poorer the definition of the bimodality. The reason for this is that the average pore size in clay soils is smaller than that in sandy soils, and consequently, individual pixels are more likely to contain both pore and particle features; therefore, when the pore size approaches the detection limit of our SEM (see section 3.3), the definition of the image is poor, as is the resultant histogram.

The micrograph is converted into a binary image; this conversion is based on the histogram where a threshold level needs to be established in order to resolve pore features from particle features. Figure 3 shows a typical histogram for a BE image in which the particle and pore spaces are represented by individual peaks. The valley between the peaks contains ambiguous pixels that need to be resolved into particles or pores. We evaluated three methods on the basis of the observation that the histogram is a superposition of two overlapping Gaussian curves. However, for finer-textured soils the representation of the histograms is complex and is not normally distributed in all cases. Since in the present study we have samples with a

Table 1. Clay, Silt, and Sand Percentages, Saturated Hydraulic Conductivity K_{sat}, Porosity Measured With Image Analysis ϕ_{SEM}, Average Pore Diameter D_{avg}, Circularity C, Roughness R, Electrical Conductivity (EC), pH, Sodium Adsorption Ratio (SAR), Pore Area A, Pore Perimeter P, Particle Density ρ_d, and Bulk Density ρ_B of Gilman Silt Loam Soil, Coachella Valley, California

Sample Number	Clay, %	Silt, %	Sand, %	K_{sat}, mm h^{-1}	ϕ_{SEM}, %	D_{avg}, μm	C	R	EC, dS m^{-1}	pH	SAR	A, μm^2	P, μm	ρ_d, g cm^3	ρ_B, g cm^3
1	9	35	56	3.18	31.1	12.9	2.97	1.06	25.4	7.25	71	444	118	2.57	1.38
2	10	39	51	2.84	35.5	12.2	3.21	1.05	25.4	7.25	71	505	116	2.55	...
3	10	50	37	2.12	29.2	10.9	3.15	1.04	25.4	7.25	71	144	58	2.71	1.38
4	16	50	34	3.24	32.4	11.4	3.23	1.03	62.7	7.36	217	498	120	2.67	1.37
5	34	61	5	0.03	26.0	10.6	3.13	1.05	62.7	7.36	217	97	57	2.71	1.35
6	26	67	7	0.05	27.3	10.6	3.01	1.05	62.7	7.36	217	81	51	2.65	1.36
7	7	21	73	19.9	34.7	11.8	3.16	1.02	1.50	7.44	3.8	887	145	2.55	1.39
8	8	16	76	17.6	32.7	11.4	3.21	1.05	1.50	7.44	3.8	385	100	2.57	...
9	4	7	89	50.3	37.6	11.8	3.39	1.02	1.50	7.44	3.8	2204	195	2.54	...
10	11	25	64	2.77	36.7	11.0	3.34	1.03	1.16	7.63	4.7	530	112	2.57	1.49
11	8	23	69	6.80	39.4	11.1	3.13	1.03	1.16	7.63	4.7	677	103	2.58	...
12	9	28	63	2.04	28.7	10.7	3.34	1.03	1.16	7.63	4.7	188	69	2.59	1.44
13	10	40	50	2.09	35.5	12.8	3.52	1.04	1.20	7.50	3.0	361	114	2.55	1.31
14	12	34	54	2.31	29.8	11.9	3.33	1.04	1.20	7.50	3.0	230	85	2.54	...
15	10	41	49	4.38	31.0	11.4	3.21	1.05	1.20	7.50	3.0	191	79	2.66	1.37
16	18	48	34	0.14	25.6	11.5	3.27	1.06	6.41	7.90	12	121	64	2.69	1.42
17	16	49	36	0.54	21.4	10.4	3.21	1.03	6.41	7.90	12	125	57	2.77	1.50
18	15	54	31	0.37	34.0	11.3	3.48	1.04	6.41	7.90	12	283	94	2.78	1.43
19	17	57	26	0.36	32.5	11.4	3.21	1.07	3.20	7.65	2.0	161	82	2.69	1.35
20	23	56	21	0.64	21.9	10.4	3.21	1.03	3.20	7.65	2.0	102	56	2.78	...
21	24	60	16	0.19	30.8	10.4	3.17	1.03	3.20	7.65	2.0	185	67	2.74	1.28
22	15	50	35	0.79	33.0	12.1	3.21	1.06	3.22	7.25	3.5	273	93	2.62	1.40
23	12	44	49	2.25	30.8	11.0	3.21	1.03	3.22	7.25	3.5	252	80	2.64	...
24	15	47	37	1.29	37.2	11.8	3.21	1.06	3.22	7.25	3.5	321	93	2.62	1.29
25	10	33	56	1.74	28.7	13.0	3.52	1.05	3.31	7.90	8.0	276	104	2.63	1.40
26	9	32	59	2.76	29.5	11.5	3.32	1.04	3.31	7.90	8.0	228	84	2.51	...
27	11	33	56	1.71	25.4	11.2	3.20	1.04	3.31	7.90	8.0	159	72	2.57	1.38
28	13	32	55	0.36	26.1	11.8	3.24	1.03	4.17	7.50	5.2	275	98	2.72	1.51
29	11	29	60	0.73	35.3	12.1	3.40	1.06	4.17	7.50	5.2	219	93	2.61	1.47
30	9	28	63	0.38	19.4	10.7	3.18	1.03	4.17	7.50	5.2	153	60	2.68	1.48
31	15	30	55	0.55	23.0	10.8	3.17	1.04	5.40	8.10	7.2	155	69	2.52	1.53
32	16	44	40	0.44	22.3	11.0	3.16	1.03	5.40	8.10	7.2	202	72	2.71	1.31
33	17	45	38	0.19	30.0	11.2	3.20	1.05	5.40	8.10	7.2	198	74	2.79	1.53
34	20	57	30	0.06	35.4	11.5	3.21	1.07	65.8	7.95	260	192	81	2.77	1.41
35	21	54	25	0.09	30.9	11.2	3.21	1.05	65.8	7.95	260	135	68	2.52	1.31
36	20	54	25	0.09	28.8	11.8	3.21	1.08	65.8	7.95	260	129	73	2.81	1.28

The ρ_B values not shown are missing values; for these samples we used the average $\rho_B = 1.43$ g cm^3.

wide range in textural composition, we will use the mathematical minimum of the histogram to establish the threshold level to convert the picture into a binary image.

3.2. Two-Point Correlation Functions

A two-point correlation function provides the probability that two points at a specific distance are located in the region of space occupied by one constituent of a two-phase material. These functions have been used extensively to calculate different properties of heterogeneous composites and provide information about the representativeness of the image. In our case the function is derived from the binary BE image and provides the probability that two pixels separated by a distance k are simultaneously in either a pore or a particle. When considering pores, if $k = 0$ (zero lag), the probability is equal to the porosity ϕ. At large distances and isotropic conditions the probability is equal to ϕ^2 [Blair et al., 1996].

Two-point correlation functions of the binary picture $f(i, j)$ with N vertical and M horizontal pixels can be calculated as the two-dimensional correlation matrix $S_2(m, n)$ expressed as

$$S_2(m, n) = \frac{1}{M_{max}N_{max}} \sum_{i=1}^{M_{max}} \sum_{j=1}^{N_{max}} f(i, j) f(i + m, j + n) \quad (1)$$

where n and m are the vertical and horizontal lags and $N_{max} = N - n$ and $M_{max} = M - m$; $f(i, j) = 1$ for pixels in pores and 0 for pixels in particles. The calculation of $S_2(m, n)$ with (1) is very inefficient for an image of significant size. A more
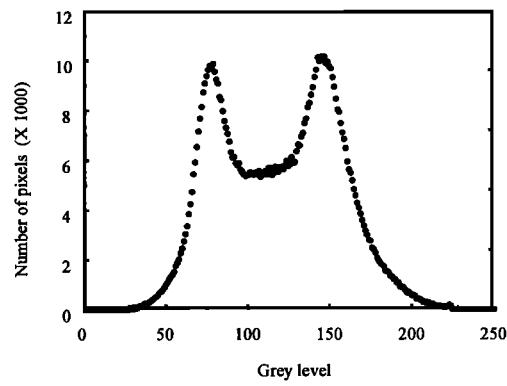

Figure 3. Histogram of the grey level of a backscattered electron image of a soil thin section.

effective analysis is to perform a two-dimensional fast Fourier transform (FFT) of the matrix $f(i, j)$ as

$$\mathbf{P}(a, b) = \text{FFT}[f(i, j)] \quad (2)$$

where a and b are indices in the complex frequency matrix $\mathbf{P}$. For each frequency a, b the correlation is performed as [*Press et al.*, 1988]

$$\mathbf{C}(a, b) = \mathbf{P}(a, b)\mathbf{P}^*(a, b) \quad (3)$$

where $\mathbf{P}^*(a, b)$ is the complex conjugate of $\mathbf{P}(a, b)$. Through the inverse FFT of $\mathbf{C}(a, b)$ we obtain the two-dimensional correlation matrix $S_2(m, n)$:

$$S_2(m, n) = \text{FFT}^{-1}[\mathbf{C}(a, b)] \quad (4)$$

The column $S_2(m = 1, n)$ and row $S_2(m, n = 1)$ are the vertical and horizontal correlation functions and provide information on the isotropy of the sample. The two-dimensional correlation matrix is summarized into a one-dimensional two-point correlation function by using the procedure outlined by *Berryman* [1998].

3.3. Determination of Optimal Magnification

Magnification is the most important factor affecting the quantification in the analysis of an image. If we measure the same parameter, for example, the hydraulic radius ($r_H = A/P$), at the same location of the same sample at different magnifications, we observe that the value of r_H varies by >50% (Figure 4). The definition of the SEM image determines the quality of the measurements, which depends on the pixel size, which, in turn, is inversely proportional to the magnification. *Berryman and Blair* [1987] also found that image magnification was important in the determination of a surface to volume ratio for use in a Kozeny-Carman model. They provide a methodology, based on a two-point correlation function (4), for finding the appropriate pixel size to use when preparing images of cross sections to be used to study water flow. The two-point correlation function that we use in this study was calculated using (2), (3), and (4). From (4) we obtain a characteristic length which is an estimate of the mean pore radius r and define the range in which this r must be in order to determine reasonable porosity measurements. *Berryman and Blair* [1987] recommend that the pixel size be ~1% of the size of an average pore radius and that

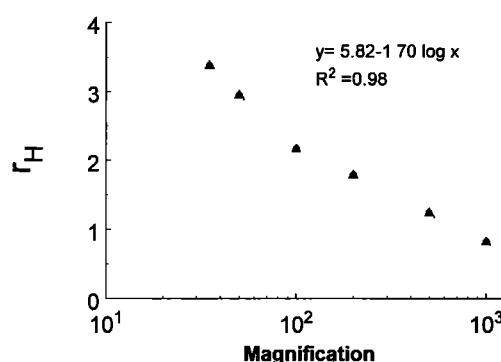$$h \geq \frac{2r}{N(1 - \phi)} \quad (5)$$

Figure 4. Hydraulic radius r_H as a function of the magnification.

Table 2. SEM Measurements for Sample 25 at Different Magnifications: Pixel Size h, Mean Pore Diameter D_m, Area A, Perimeter P, and Porosity Φ_{SEM}

Magnification, ×	h, μm	D_m, Pixels	Φ_{SEM}	Φ_{TCF}	A, μm^2	P, μm
35	2.691	12.87	26.7	29.2	442	131
50	1.884	8.25	30.3	29.8	415	141
100	0.942	5.70	30.8	33.4	171	79.5
200	0.471	4.32	32.7	34.0	64.5	36.0
500	0.188	2.63	32.4	34.9	26.3	21.3
1000	0.094	2.05	38.4	50.8	13.8	16.9

Also given is porosity calculated by the two-point correlation function Φ_{TCF}.

where h is the pixel size and N is the number of pixels along each side of a square image. Berryman and Blair emphasize that there is no advantage to increasing magnification if the average pore diameter is larger than a quarter of the width of the digitized image and recommend having ~100 grains in the picture for measurement of porosity.

Table 2 shows the r, pixel size, magnification, and porosity values calculated with (4) for sample 25 in Table 1. Calculating the right-hand side of (5) for an image of 1000 pixels, we obtain the value $h \geq 0.1$. When this value is combined with the requirement of having at least 100 features in the picture, we determine that magnifications within the acceptable range are 200–35×. A magnification of 50× was chosen as a compromise between a bigger sampling size and a satisfactory detection limit. The detection limit is 7 μm at 50× magnification and the resolution chosen (1024 × 800 pixels).

4. Prediction of Saturated Hydraulic Conductivity

4.1. Kozeny-Carman Equation

There are a number of methods available to relate SEM data to saturated hydraulic conductivity. Empirical relationships such as fractal analysis or neural network have been used in the literature, however, it is more informative to include as many physical relationships as possible. The model traditionally used to predict K_{sat} is the Kozeny-Carman equation, but different approaches can be taken; for example, *Ahuja et al.* [1989], using macroscopic measurements, proposed

$$K_{sat} = Q \phi_e^B \quad (6)$$

where ϕ_e is the effective porosity, defined as the difference between the soil water retention at saturation and the water retention at 33 kPa. The parameters Q and B are fitting parameters, and they depend on the calibration data set [cf. *Gimenez et al.*, 1997].

A different approach is that followed by *Blair et al.* [1996], who used the Kozeny-Carman relation to interpret data derived from SEM images of sandstones. They used the expression:

$$K_{sat} = C \phi r_H^2 \quad (7)$$

where ϕ is the porosity and r_H is the hydraulic radius. The constant C can be defined by using the Poiseuille equation:

$$C = \frac{g}{v G \tau^2} \quad (8)$$

where C incorporates effects of acceleration of gravity g, pore geometry G, pore tortuosity τ, and the kinematic viscosity of the pore water v. C has the largest value for parallel, circular pores ($G = 8$, $\tau = 1$). However, for real soils it is most likely that the assumption of parallel, circular pores will lead to a considerable overestimation of the saturated hydraulic conductivity. Furthermore, pore connectivity is an additional factor that may play a role in soils.

In this study we will calculate the hydraulic radius r_H from total pore area A and total pore perimeter P

$$r_H = A/P \qquad (9)$$

This definition will include noncircularity and roughness of the pores with both A and P measured directly with the SEM. For equal areas the hydraulic radius will become smaller when the noncircularity or pore roughness increases. The porosity ϕ in (7) is also measured from the SEM image analysis (ϕ_{SEM}). The constant C in (7) is split into a constant part $C_1 = g/v$ (equal to 8.46×10^9) and a nonconstant part F that lumps effects of pore tortuosity, geometry, and connectivity. This factor F is a soil-specific reduction factor and is analogous to the "formation factor" that is used in many hydrological studies of rock strata [e.g., *Adler et al.*, 1992; *Berryman and Blair*, 1987; *Blair et al.*, 1996]. To remove a bias toward high hydraulic conductivities, we express (7) in a logarithmic form:

$$\log (K_{sat}) = \log (C_1) + \log (F) + \log [\phi_{SEM}(A/P)^2] \qquad (10)$$

Equation (10) will be used in two different ways. First, we assume that the concept of straight cylindrical pores applies to our samples and predict K_{sat} with (10) using measured ϕ_{SEM}, A, and P with $C = 8.46 \times 10^9$ and $F = \frac{1}{8}$ ($G = 8$ and $t = 1$). Second, we will use measured K_{sat} and solve (10) for F. This approach is necessary because we cannot account for the individual contributions of pore tortuosity, geometry, and connectivity. Subsequently, we will investigate the relationship of F with microscopic properties (ϕ_{SEM}, A, P, pore roughness, and circularity), macroscopic properties (clay and sand percentage and bulk density), and soil chemical properties (pH, EC, and SAR). We will use a combination of a neural network and the bootstrap method to find a predictive model for F.

4.2. Neural Network and Bootstrap Analysis

A big advantage of neural networks over traditional regression techniques is that they do not need an a priori selection of the underlying model expressions (linear, exponential, etc). The term "neural network" represents a large collection of numerical techniques that resemble biological neural systems. Many different types of neural network exist, each with a particular range of applications [*Hecht-Nielsen*, 1991; *Haykin*, 1994].

In this study we used the most common type of three-layer feed-forward networks as also used by *Pachepsky et al.* [1996], *Schaap and Bouten* [1996], and others. This type of neural network can be seen as a "universal function approximator" because it can estimate any continuous nonlinear function with a desired degree of accuracy. Because they need no a priori model specifications, neural networks are capable of finding complex relationships between input data (in our case, for example, clay percentage, bulk density, pore roughness, and pH) and output data (the factor F). An iterative calibration procedure based on the Levenberg-Marquardt algorithm [*Marquardt*, 1963; *Demuth and Beale*, 1992] was used to find these

relationships through minimization of the following objective function:

$$Ob = \sum_{i=1}^{36} (F - F')^2 \qquad (11)$$

where the number of samples is 36 and F and F' are calculated with (10) and the neural network–predicted F factors.

Typically, usage of neural networks leads to black box models in which the flow of information can be difficult to track. Furthermore, neural networks tend to be sensitive to nonlinear instability and overfitting [*Hsieh and Tang*, 1998]. Nonlinear instability stems from the possibility that many local minima may exist on the surface of the objective function. Overfitting results from the function approximation of the neural networks, which, without further precaution, may lead to the undesired inclusion of noise and artifacts into the model. A model that is overfitted will give very good predictions for its calibration data set but poor predictions for independent data. These problems can largely be overcome by repeatedly restarting the neural network calibration with different initial coefficients, by limiting the number of iterations [*Schaap et al.*, 1999], and by validating the neural network model on independent data.

The bootstrap method [*Efron and Tibshirani*, 1993] was used to generate uncertainty estimates of predicted F and to generate independent calibration and validation data sets for the neural network calibration. Bootstrap theory assumes that multiple alternative realizations of the population can be simulated from the single data set that is available. By calibrating the model on these alternative data sets, different predictions result, leading to uncertainty about the true model. The alternative data sets have the same size as the original data set and are created by random resampling with replacement. Therefore, in a data set of n samples, each sample has a chance of $1 - [(n - 1)/n]^n$ of being selected once or multiple times. Because some samples are selected more than once, each alternative data set contains ~64% (for $n = 36$) of the original data. Neural networks were calibrated on each of these data sets. The bootstrap method was combined with the TRAINLM routine of the neural network toolbox [*Demuth and Beale*, 1992] of the MATLAB® package (version 4.0, MathWorks Inc., Natick, Massachusetts). The neural network code was modified to avoid local minima in the objective function. We used 60 alternative data sets, leading to 60 neural network models; the average of the 60 predictions was the true F, and the corresponding standard deviation gave the uncertainty in F.

5. Results and Discussion

5.1. Microscopic Method

The use of microscopic techniques to describe the pore space in a soil has the advantage that we can consider pore properties that are otherwise unavailable. However, special precautions need to be taken if the objective is to relate those properties with macroscopic events. As a preliminary evaluation of the consistency of our microscopic technique, we compared the porosity of our samples obtained by the image analysis from the micrographs (ϕ_{SEM}) with the porosity calculated using

$$\phi = 1 - \frac{\rho_B}{\rho_d} \qquad (12)$$

where ρ_B is the bulk density and ρ_d is the soil particle density. Figure 5 shows that the porosity measured with SEM systematically underestimates the total calculated porosity (12) by ~ 0.15 cm^3/cm^3. Most likely this underestimation is due to unresolved pores <7 μm. To explain the difference between ϕ_{SEM} and calculated porosities, we derived the pore volume contained in pores <7 μm using a pedotransfer function [Schaap et al., 1998]. By using the capillary law it follows that a pore of 7 μm in diameter is equivalent to a pressure head of 43 kPa. After adding the porosity associated with the water content at 43 kPa pressure head to ϕ_{SEM} we found a much better agreement with the calculated values (Figure 5). Even though the pedotransfer function provides predictions with an implicit error, we estimate that our porosity measurements are comparable to those obtained conventionally.

The two-point correlation function calculations for all our samples indicated that the magnification used (50×) provides acceptable resolution. As mentioned earlier, the resolution of the images is related to the clay content of the sample. Figure 6 shows the micrographs and the correlation functions for samples 5 and 9 (Table 2), whose clay content covers the range of our samples. The correlation functions show that the characteristic length (where the function crosses the ϕ_{SEM}^2) is much smaller than the size of the micrograph (1880 μm). The accumulation of 10 micrographs per sample ensures a good representation of the microfeatures.

From these results we conclude that a magnification of 50× leads to a sufficient resolution of relevant small pores, as well as a good spatial coverage. Furthermore, the magnification will also lead to an exclusion of features that are too small to have an effect on saturated hydraulic conductivity [Ahuja et al., 1989; Berryman and Blair, 1987].

5.2. Prediction of Saturated Hydraulic Conductivity

As we reviewed in section 4.1, there are different possible expressions of the Kozeny-Carman equation. We tested the relationship proposed by Ahuja et al. [1989] to predict K_{sat} in our soils samples. When we substituted ϕ_{SEM} in (6) and used linear regression based on log (K_{sat}) values, we obtained $Q = 304$ and $B = 4.07$ for our data set. However, the prediction of K_{sat} versus measured K_{sat} had a very low correlation coefficient (0.17, Table 3). The lack of correlation between mea-

A

B

C

Figure 6. Binary micrographs for (a) sample 5, (b) sample 9, and (c) two-point correlation function for samples 5 and 9.
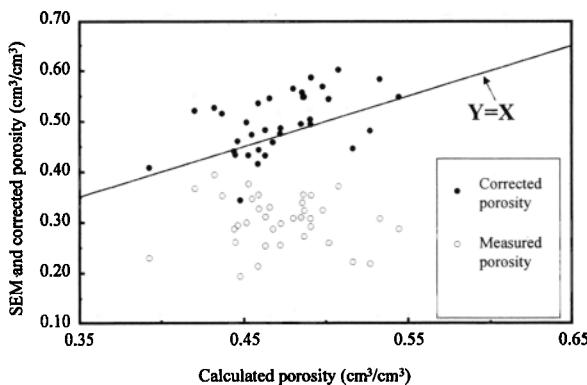

Figure 5. Porosity of the 36 samples measured with the scanning electron microscope before and after the correction for the detection limit versus the calculated porosity.
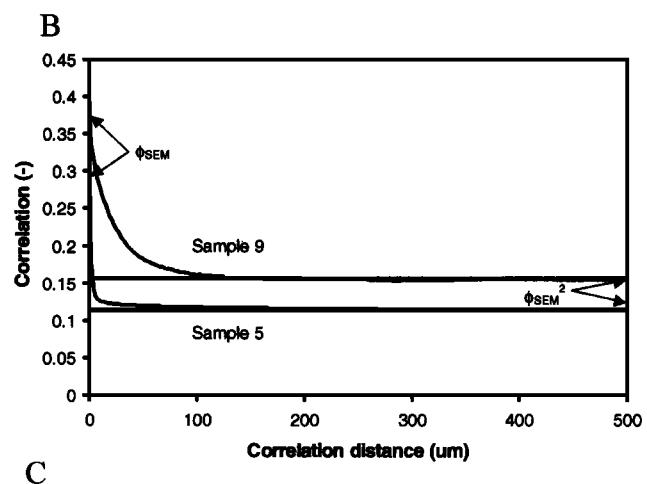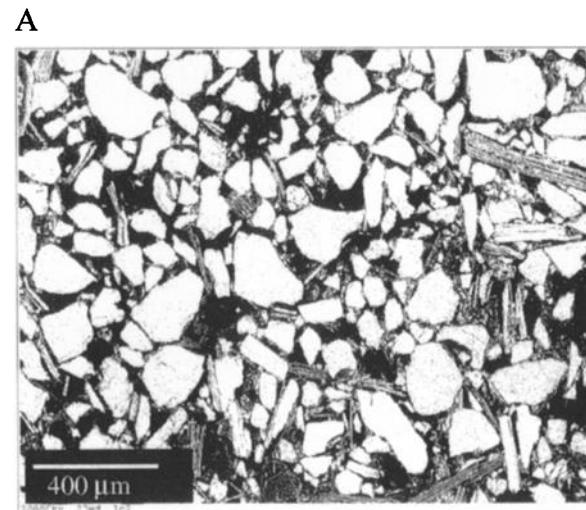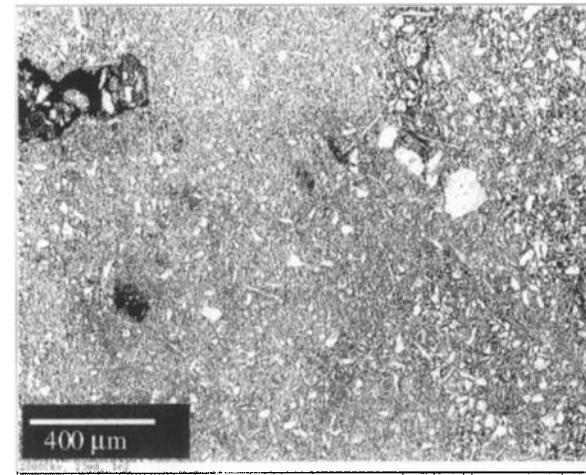
sured and predicted K_{sat} indicates that the version of the Kozeny-Carman equation by Ahuja et al. [1989] cannot be reliably used for our data. The use of a higher effective porosity (the cutoff pore size in our case corresponds to 43 kPa instead of the 33 kPa proposed by Ahuja et al. [1989]) does not justify the poor prediction and indicates that fitting parameters

Table 3. Correlation Coefficient R^2 of the Regression Between the Measured K_{sat} and the Predicted K_{sat} Using Different Models

Model	K_{sat} Predictions (R^2)
$K_{sat} = A\phi_{SEM}^B$	0.17
Neural network	
Input SSC	0.71
Input SSC + ρ_B	0.75
Input SSC + ρ_B + SAR + pH + EC	0.77
$K_{sat} = FC\phi_{SEM}r_H^2$ (F obtained with neural network and input C + ρ_B + pH + R)	0.91

The first model is the Kozeny-Carman equation used by *Ahuja et al.* [1989] and *Gimenez et al.* [1999], the second model is direct regressions of different variables with measured K_{sat} using neural network analysis, and the third model is the Kozeny-Carman equation used in the present study, for which the factor F was predicted using neural network analysis. SSC, sand, silt, and clay; ρ_B, bulk density; SAR, sodium absorption ratio; EC, electrical conductivity; and R, roughness.

without a physically based conceptual model is not adequate for our soil samples.

The conceptual model needs to account for the difference between an ideal system of an array of parallel cylindrical pores and the intricate distribution of pores in a natural soil. The pore space is the result of the rearrangement of particles in structural units called aggregates. These aggregates are the result of a series of attractive and repulsive forces of electrostatic and physical origin. We introduced the factor F, which is a combination of tortuosity, connectivity, and microscopic parameters. F was best predicted when clay percentage, ρ_B, roughness, and pH were used as predictors using neural network analysis. Clay percentage was the single variable with the highest contribution ($R^2 = 0.60$ alone) to the prediction of F.

The prediction of the formation factor did not require salinity or sodicity parameters. However, it is well known, from dispersion or aggregate stability tests, that salinity and sodicity have an important role in the dispersion and flocculation of clay particles [*Goldberg and Forster*, 1990; *Hesterberg and Page*, 1993; *Lebron and Suarez*, 1992; *Kretzschmar et al.*, 1993], in the hydraulic conductivity [*Suarez et al.*, 1984], and, consequently, in the rearrangement of the clay domains [*Shainberg and Otho*, 1968; *Quirk and Aylmore*, 1971; *Lebron et al.*, 1993].

However, when EC or SAR were incorporated into the neural network analysis to predict F, there was no significant increase in the correlation coefficient. From Table 1 we observe that in this particular sample population, EC and SAR

are not independent variables and show a strong linear correlation. It is known that EC and SAR have opposite effects in the development of the double layer and, consequently, in the flocculate/dispersed state of the clay particles; while increases of EC favor flocculation, increases of SAR promote dispersion. This direct relation of sodicity and salinity is common in arid zones where the majority of the salts are sodic.

The values of pH were not lineally correlated with the chemical parameters, and the prediction of F was improved after the incorporation of pH in the neural network analysis. The pH has been widely neglected in transport modeling efforts, but *Suarez et al.* [1984] showed that at high pH an increase in 1 unit of pH can cause a decrease in K_{sat} of 1 order of magnitude; this decrease is equivalent to an increase of 20 SAR units. The importance of the pH is related to the sign of the variable charges, generally located at the edges of the clays, in the iron and aluminum oxides, and in the organic matter. Above the point of zero charge (PZC) the variable charges are negative; consequently, the net electrostatic charges in the soil particles are negative, and repulsive forces occur among the particles. When the pH of the soil is below the PZC, the edges are positively charged, providing the possibility for the electrostatic forces to form bonds. There is a considerable variability of PZC values for the different components of a soil, but in general, we can consider that there is a surface charge reversal around pH 7-9. When the predicted F values were calculated in the range of 5-25% clay at constant bulk density and roughness and variable pH (Figure 7), the F factor decreased when the pH increased and decreased when the clay content increased, in agreement with our previous observations.

If we keep pH and ρ_B constant and change roughness, we predict a decrease in F values when the roughness increases (Figure 8); this agrees with the results of *Berryman and Blair* [1987]. Berryman and Blair found that for the Kozeny-Carman model for fluid flow in porous media it is not the absolute specific surface area but rather the roughness of the pore walls relative to the mean pore size that influences the permeability. When roughness and pH were constant, the F factor decreased when the ρ_B increased (Figure 9); this may due to the fact that increases in ρ_B are associated with an increase in tortuosity, which is known to reduce the K_{sat}.

The predicted K_{sat}, taking into account the F factor in the Kozeny-Carman equation, shows a very good agreement with the measured K_{sat} (Figure 10) with a regression coefficient of 0.91. The predictions were highly improved by the use of r_H,
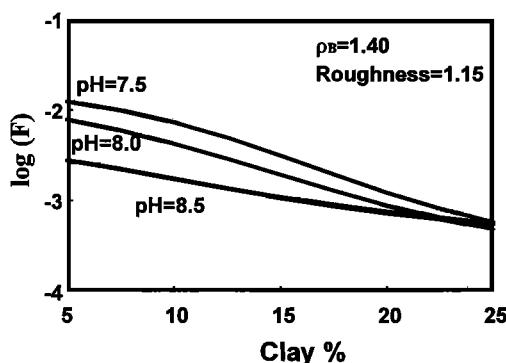

Figure 7. Predicted values of the log of the formation factor F at different pH values in the range of 5-25% clay content.
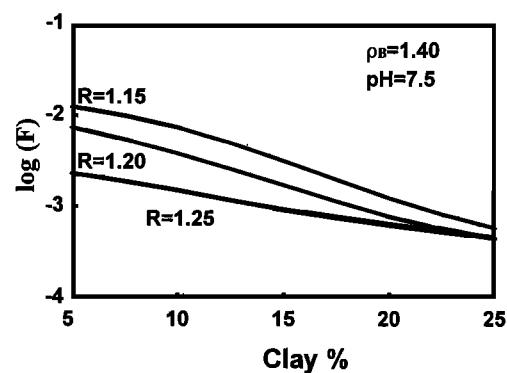

Figure 8. Predicted values of the log of the formation factor F at different roughness values R in the range of 5-25% clay content.

which includes area and perimeter of the pores directly measured in the image. This information cannot be obtained by any other methodology (that is not image based); r_H is traditionally inferred through indirect methods with a series of assumptions such as the absence of coulombic forces among particles and specific geometry. As a corroboration of the need for a physically based model we applied neural network analysis directly to our data in an attempt to empirically predict K_{sat} without using the Kozeny-Carman equation or microscopic parameters. We used different variables with relative success; Table 3 shows that we were able to predict K_{sat} with a correlation coefficient range of 0.71–0.77, but we did not reach the quality of the predictions made with (10).

Despite the excellent correlation between the calculated and measured K_{sat} there are many unresolved problems that prevent the use of the F values predicted in this study for soils from nonarid regions. It is for that reason that more studies are needed to quantify the effects of EC and sodicity on pore geometry and pore distribution. The individual contribution of each one of the chemical parameters will provide the essential information to develop a conceptual model for the prediction of hydraulic properties. Also, measurements of tortuosity and connectivity with independent methodologies, such as time domain reflectometry or acoustic techniques, will provide the necessary information for a more physically based model. The results of such measurements will be addressed in future experiments.

6. Conclusions

The proposed methodology provides a quantification of parameters of the pores at a scale not possible to measure with macroscopic techniques. Microscopic measurements of the pore space have been shown to be in good agreement with the bulk porosity of the soils calculated with traditional methods.

The use of macroscopic parameters in the Kozeny-Carman equation yielded poor predictions of K_{sat}; however, when microscopic information of the pores was included, we were able to predict K_{sat} with $R^2 = 0.91$ for the same soil population. Roughness and pH have been shown to be effective for the prediction of the formation factor F of the soils examined.

The methodology proposed in the present study shows very promising results for improving prediction of hydraulic properties by incorporating a more realistic quantification of the porous media, but more studies are needed to develop a con-
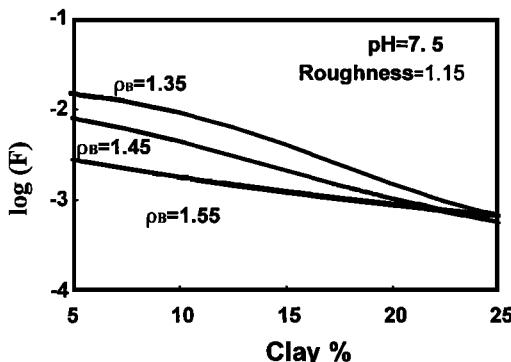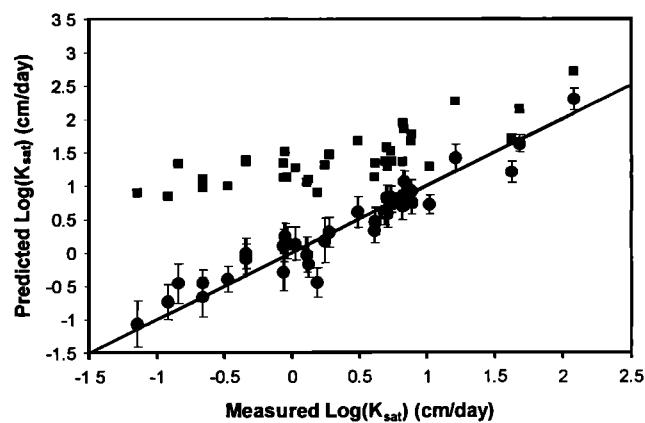
Figure 10. Predicted K_{sat} using the Kozeny-Carman equation with (circles) and without (squares) the formation factor versus the measured K_{sat}.

ceptual model of the effect of the chemical properties on the pore space and pore distribution in the soil.

References

Adler, P. M., C. G. Jacquin, and J. F. Thovert, The formation factor of reconstructed porous media, *Water Resour. Res.*, *28*, 1571–1576, 1992.

Ahuja, L. R., D. K. Cassel, R. R. Bruce, and B. B. Barnes, Evaluation of spatial distribution of hydraulic conductivity using effective porosity data, *Soil Sci.*, *148*, 404–411, 1989.

Berryman, J. G., Measurement of spatial correlation functions using image processing techniques, *J. Appl. Phys.*, *57*, 2374–2384, 1998.

Berryman, J. G., and S. C. Blair, Kozeny-Carman relations and image processing methods for estimating Darcy's constant, *J. Appl. Phys.*, *62*, 2221–2228, 1987.

Blair, S. C., P. A. Berge, and J. G. Berryman, Using two-point correlation functions to characterize microgeometry and estimate permeabilities of sandstones and porous glass, *J. Geophys. Res.*, *101*, 20,359–20,375, 1996.

Blake, G. R., and K. H. Hartge, Particle density, in *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods, Agron. Monogr. 9*, 2nd ed., edited by A. Klute, pp. 377–382, Am. Soc. Agron., Madison, Wisc., 1986.

Bruand, A., I. Cousin, B. Nicoullaud, O. Duval, and J. C. Begon, Backscattered electron scanning images of soil porosity for analyzing soil compaction around roots, *Soil Sci. Soc. Am. J.*, *60*, 895–901, 1996.

Chretien, J., and E. B. A. Bisdom, The development of soil porosity in experimental sandy soils with clay admixtures as examined by Quantimet 720 from BESI and by other techniques, *Geoderma*, *30*, 285–302, 1983.

Demuth, H., and M. Beale, Neural network toolbox manual, MathWorks Inc., Natick, Mass., 1992.

Efron, B., and Tibshirani, R. J., *An Introduction to the Bootstrap*, Monogr. Stat. Appl. Probab., Chapman and Hall, New York, 1993.

Ehrlich, R., S. J. Crabtree, K. O. Horkowitz, and J. P. Horkowitz, Petrography and reservoir physics, I, Objective classification of reservoir porosity, *AAPG Bull.*, *75*, 1547–1562, 1991.

Friedman, G. M., Determination of sieve-size distribution from thin section data for sedimentary petrological studies, *J. Geol.*, *66*, 394–416, 1958.

Friedman, G. M., Comparison of moment measures for sieving and thin section data for sedimentary petrological studies, *J. Sediment. Petrol.*, *32*, 15–25, 1962.

Gee, G. W., and J. W. Bauder, Particle size analysis, in *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods, Agron. Monogr.*

Figure 9. Predicted values of the log of the formation factor F at different bulk density values ρ_B in the range of 5–25% clay content.

9, 2nd ed., edited by A. Klute, pp. 383–412, Am. Soc. Agron., Madison, Wisc., 1986.

Gimenez, D., R. R. Allmaras, D. R. Huggins, and E. A. Nater, Prediction of the saturated hydraulic conductivity-porosity dependence using fractals, Soil Sci. Soc. Am. J., 61, 1285–1292, 1997.

Gimenez, D., W. J. Rawls, Y. A. Pachepsky, and J. P. C. Watt, Characterization of soil structure in relation to saturated hydraulic conductivity, in Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media, edited by M. Th. van Genuchten et al., pp. 1019–1028, Univ. of Calif. Press, Berkeley, in press, 1999.

Goldberg, S., and H. S. Forster, Flocculation of reference clays and arid-zone soil clays, Soil Sci. Soc. Am. J., 54, 714–718, 1990.

Haykin, S., Neural Networks, A Comprehensive Foundation, 1st ed., Macmillan, Indianapolis, Ind., 1994.

Hecht-Nielsen, R., Neurocomputing, 1st ed., Addison-Wesley, Reading, Mass., 1991.

Hesterberg, D., and A. L. Page, Critical coagulation concentration of sodium and potassium illite as affected by pH, Soil Sci. Soc. Am. J., 54, 735–739, 1993.

Hsieh, W. W., and B. Tang, Applying neural network models to prediction and data analysis in meteorology and oceanography, Bull. Am. Meteorol. Soc., 79, 1855–1870, 1998.

Kretzschmar, R., W. P. Robarge, and S. B. Weed, Flocculation of kaolinitic soil clays: Effects of humic substances and iron oxides, Soil Sci. Soc. Am. J., 57, 1277–1283, 1993.

Lebron, I., and D. L. Suarez, Variations in soil stability within and among soil types, Soil Sci. Soc. Am. J., 56, 1412–1421, 1992.

Lebron, I., D. L. Suarez, C. Amrhein, and J. E. Strong, Size of mica domains and distribution of the adsorbed Na-Ca ions, Clays Clay Miner., 41, 380–388, 1993.

Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math., 11, 431–441, 1963.

Pachepsky, Y. A., D. Timlin, and G. Varallyay, Artificial neural networks to estimate soil water retention from easily measurable data, Soil Sci. Soc. Am. J., 60, 727–733, 1996.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes in C, 1st ed., Cambridge Univ. Press, New York, 1988.

Quirk, J. P., and L. A. G. Aylmore, Domains and quasi-crystalline regions in clay systems, Soil Sci. Soc. Am. Proc., 35, 652–654, 1971.

Schaap, M. G., and W. Bouten, Modelling water retention curves of sandy soils using neural networks, Water Resour. Res., 32, 3033–3040, 1996.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten, Neural network analysis for hierarchical prediction of soil hydraulic properties, Soil Sci. Soc. Am. J., 62, 847–855, 1998.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten, A bootstrap-neural network approach to predict soil hydraulic parameters, in Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media, edited by M. Th. van Genuchten, F. J. Leij, and L. Wu, pp. 1237–1250, Univ. of Calif. Press, Berkeley, in press, 1999.

Schoonderbeek, D., F. Thiel, and E. B. A. Bisdom, Quantimet 720 analysis of porosities in backscattered electron scanning images made with different phototechniques, in Submicroscopic Studies of Soils, edited by E. B. A. Bisdom and J. Ducloux, pp. 271–275, Elsevier, New York, 1983.

Shainberg, I., and H. Otoh, Size and shape of montmorillonite particles saturated with Na/Ca ions inferred from viscosity and optical measurements, Isr. J. Chem., 6, 251–259, 1968.

Suarez, D. L., J. D. Rhoades, R. Lavado, and C. M. Grieve, Effect of pH on saturated hydraulic conductivity and soil dispersion, Soil Sci. Soc. Am. J., 48, 50–55, 1984.

Tuller, M., D. Or, and L. M. Dudley, Adsorption and capillary condensation in porous media: Liquid retention and interfacial configurations in angular pores, Water Resour. Res., 35, 1949–1964, 1999.

Vogel, H. J., Morphological determination of pore connectivity as a function of pore size using serial sections, Eur. J. Soil Sci., 48, 365–377, 1997.

I. Lebron, M. G. Schaap, and D. L. Suarez, U.S. Salinity Laboratory, USDA-ARS, 450 W. Big Springs Road, Riverside, CA 92507. (ilebron@ussl.ars.usda.gov)