"The **more** opportunities there are to get **students involved**, the more you will **encourage** previously unreached and **unrepresented groups** to join the Earth and Space science community."

Ryan Haupt
Research Fellow,
Smithsonian Museum
of Natural History
2015 Student Travel
Grant Recipient

**Support the next generation of Earth and space scientists.
Donate to the Austin Student Travel Grant Challenge.**

AGU100 ADVANCING EARTH AND SPACE SCIENCE

agu.org/austin | #AGU100

Developing Joint Probability Distributions of Soil Water Retention Characteristics

ROBERT F. CARSEL

Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, Georgia

RUDOLPH S. PARRISH

Computer Sciences Corporation, U.S. Environmental Protection Agency, Athens, Georgia

A method is presented for developing probability density functions for parameters of soil moisture relationships of capillary head $[h(\theta)]$ and hydraulic conductivity $[K(\theta)]$. These soil moisture parameters are required for the assessment of water flow and solute transport in unsaturated media. The method employs a statistical multiple regression equation proposed in the literature for estimating $[h(\theta)]$ or $[K(\theta)]$ relationships using the soil saturated water content and the percentages of sand and clay. In the absence of known statistical distributions for either $[h(\theta)]$ or $[K(\theta)]$ relationships, the method facilitates modeling by providing variability estimates that can be used to examine the uncertainty associated with water flow or solute transport in unsaturated media.

INTRODUCTION

Assessments of groundwater contamination from chemical waste disposal or agricultural chemical application invariably include evaluation of chemical transport through the unsaturated zone. For transforming or degrading chemicals, the magnitude of contamination depends on the residence time in the unsaturated zone. The residence time is dependent on chemical and soil characteristics and meteorologic conditions. The movement of hazardous wastes or pesticides is inherently affected by soil characteristics and the associated spatial variability occurring within and among individual waste disposal sites or agricultural use areas.

The soils literature contains numerous assessments documenting the variability associated with textural and hydraulic characteristics of soils [e.g., *Jury*, 1982; *Nielsen et al.*, 1973]. The coefficient of variation (CV) is often used to represent the magnitude of variability. The CV often is found to be highest for soil hydraulic properties (e.g., hydraulic conductivity) and lowest for textural properties such as bulk density and total porosity [e.g., *Sharma and Rogowski*, 1983; *Warrick and Nielsen et al.*, 1980]. Variations in soil characteristics can contribute considerable uncertainty [e.g., *Bresler and Dagan*, 1981; *Jury*, 1982] to assessments of solute transport and groundwater contamination.

Traditionally, mathematical models have been used to evaluate the uncertainty of predicted chemical movement in the unsaturated zone. *Cox and Baybutt* [1981] have described five different modeling methods for conducting uncertainty analyses. The choice of any one method depends upon the model (or models) selected and analysis objectives. The widely used Monte Carlo procedure is suitable for developing uncertainty analyses of solute transport. These analyses make use of randomly generated time series to produce frequency distributions. Frequency distributions can be used to assess groundwater contamination by expressing the uncertainty as a probability of occurrence. Such assessments may provide estimates of various percentiles of the predicted unit solute loadings (mass per unit area) to groundwater.

Monte Carlo numerical simulation methods require probability density functions of model input parameters and, in some cases, correlations among parameters. In a typical Monte Carlo run, for example, values of the various parameters are generated randomly from hypothesized or inferred distributions. *Carsel et al.* [1988] examined uncertainty of the leaching potential of the pesticide aldicarb through a Monte Carlo simulation. Estimated distributions of field capacity and wilting point were used to characterize input parameters for the PRZM model [*Carsel et al.*, 1988]. The potential for leaching below selected depths was expressed in the form of cumulative probability distributions.

Monte Carlo techniques that are used to evaluate uncertainty of solute transport require probability distributions for hydraulic parameters that affect water-solute movement in soil. Unfortunately, such distributional and correlational information often is lacking or is not well-established. However, these obstacles have been greatly reduced by development of estimation techniques for many of the hydraulic parameters required by solute transport models [e.g., *Rawls et al.*, 1982; *Rawls and Brakensiek*, 1985; *El-Kadi*, 1981]. Application of these estimation methods provides a basis upon which associated probability distributions of model input parameters can be inferred.

Fundamental to this approach is the need to establish good approximations to empirical distributions for many parameters in several soil classifications. A family of statistical distributions can be used advantageously to provide a commonality of form that permits correlations to be incorporated. The family of distributions known as the Johnson system [*Johnson and Kotz*, 1970; *Johnson*, 1987] was used here for this purpose. This system is rich in variety of form and is especially useful for data fitting, particularly where good approximations to many empirical distributions are needed. It provides a significant advantage over alternatives by producing, after appropriate variable transformations, a set of normally distributed variables.

As part of this work, probability density functions were developed for soil-saturated hydraulic conductivity and other hydraulic parameters. In addition, joint multivariate density functions that incorporated correlations among these variables were developed for various soil textural classes. Where

Term	ln (KS)	Θr	ln (α⁻¹)	ln (N-1)
(Constant)	-8.96847	-0.0182482	5.3396738	-0.7842831
S	-	0.0087269	-	0.0177544
C	-0.028212	0.00513488	0.1845038	-
Θ_s	19.52348	0.02939286	-2.48394546	-1.062498
S^2	0.00018107	-	-	-0.00005304
C^2	-0.0094125	-0.0015395	-0.00213853	-0.00273493
Θ_s^2	-8.395215	-	-	1.11134946
SC	-	-	-	-
$S\Theta_s$	0.077718	-0.0010827	-0.0435649	-0.03088295
$C\Theta_s$	-	-	-0.61745089	-
S^2C	0.0000173	-	-0.00001282	-0.00000235
$C^2\Theta_s$	0.02733	0.0030703	0.00895359	0.00798746
$S^2\Theta_s$	0.001434	-	-0.0072472	-
SC_2	-0.0000035	-	0.0000054	-
$C\Theta_s^2$	-	-0.0023584	0.50028060	-0.00674491
$S^2\Theta_s^2$	-0.00298	-	0.00143598	0.00026587
$C^2\Theta_s^2$	-0.019492	-0.0018233	-0.00855375	-0.00610522

S = percent sand (5<S<70)
C = percent clay (5<C<60)
Θ_s = total saturated water content, cm³ cm⁻³
KS = saturated hydraulic conductivity, cm hr⁻¹
Θ_r = residual water content, cm³ cm⁻³
α = empirical constant, cm⁻¹
N = empirical constant

General regression model:

$$f(S,C,\Theta_s) = [b_0 + b_1S + b_2C + b_3\Theta_s + b_{11}S^2 + b_{22}C^2 + b_{33}\Theta_s^2$$
$$+ b_{12}SC + b_{13}S\Theta_s + b_{23}C\Theta_s$$
$$+ b_{112}S^2C + b_{223}C^2\Theta_s + b_{113}S^2\Theta_s + b_{122}SC^2$$
$$+ b_{233}C\Theta_s^2 + b_{1133}S^2\Theta_s^2 + b_{2233}C^2\Theta_s^2]$$

Fig. 1. Multiple regression model and coefficients developed by *Rawls and Brakensiek* [1985] to estimate selected soil water retention characteristics.

input variables are correlated, a properly formulated joint statistical distribution permits combinations of values to be more appropriately represented, from a frequency standpoint, in the simulation. The presence of correlations implies that some combinations of values are more probable or less probable than they otherwise would be under an assumption of independence. A joint distribution serves to better represent the relative frequencies of the variables under study. A multivariate approach also provides for variance reduction and increased resolution in the sense that the effect of a smaller change in the system can be evaluated. *El-Kadi* [1987] considered parameter correlation in relation to infiltration. He concluded that when correlation was accommodated, uncertainty was reduced by one third for the cases considered.

Correlations among input variables can be easily incorporated into a multivariate normal distribution model, provided marginal distributions of the individual variables are them-

selves normally distributed. Generally, however, if the joint distribution is not normal, it may be more difficult or impossible to include the correlation structure in the distribution model. Therefore it was a goal of this research to identify, through appropriate choices of variable transformations, the associated probability distributions as members of the Johnson family that best fitted the empirical frequency distributions. Then, estimates of covariances could be used to produce a multivariate normal distribution model that embodied all of the distributional information (for these parameters) needed for subsequent Monte Carlo modeling or other simulation studies.

DATA AND PROCEDURES

Evaluation of Soil Data

The complexity and extreme variability of soil at the scale of the primary particle can be bypassed by measuring hydro-
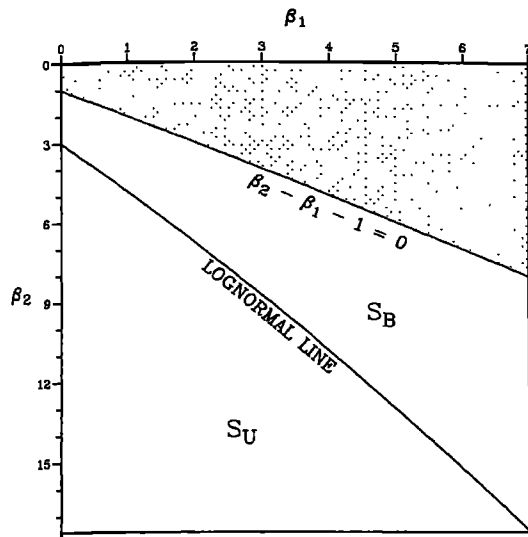
Fig. 2. *Johnson* [1987] system of distributions in relation to third and fourth standard moments.

logical properties at much larger scales [*Sharma and Rogowski*, 1983]. Various combinations are possible for producing a sample population of sufficient size to describe the variability that may be expected among soils. For example, *Baes and Sharp* [1983] reduced the apparent variability of soil bulk density estimates by grouping soils into five textural types (silt loams, clays and clay loams, sandy loams, gravelly silt loams, and loams) reported by *Holtan et al.* [1968] and *Free et al.* [1940]. The variability of these characteristics is a product of both the inherent spatial variability of the continuum and their assignment to categories.

The soil water characteristic [$h(\Theta)$] and hydraulic conductivity [$K(\Theta)$] functions are essential to the application of soil water flow theory and solute transport. Experimental methods [e.g., *Hillel*, 1982] for determining these curves often are time consuming and tedious. Thus simplified approaches for estimating the hydraulic properties of soils are quite useful, especially for nonpoint source problems. *Rawls and Brakensiek* [1985] have demonstrated a method for computing saturated hydraulic conductivity from soil-saturated water content, sand content, and clay content. Their analysis also indicated that these characteristics can be used to estimate the parameters required by several water retention models [e.g., *Brooks and Corey*, 1964; *Campbell*, 1974; *van Genuchten*, 1976].

The *van Genuchten* [1976] model is widely used for predicting soil water content as a function of pressure head. This model is generally expressed as

$$\Theta = \Theta_r + \frac{(\Theta_s - \Theta_r)}{[1 + (\alpha h)^N]^M} \qquad (1)$$

where

Θ water content;
Θ_r residual water content;
Θ_s total saturated water content;
α empirical constant, cm^{-1};
N empirical constant;
M empirical constant;
h capillary head, cm.

Also, where M is related to N as follows:

$$M = 1 - 1/N$$

Hydraulic conductivity can be represented by

$$\frac{K(\Theta)}{K_s} = \left\{\frac{\Theta - \Theta_r}{\Theta_s - \Theta_r}\right\}^{1/2}\left\{1 - \left[1 - \left(\frac{\Theta - \Theta_r}{\Theta_s - \Theta_r}\right)^{1/M}\right]^M\right\}^2 \qquad (2)$$

where $K(\Theta)$ is the hydraulic conductivity for a given water content (centimeters per hour) and K_s is the saturated hydraulic conductivity (centimeters per hour). Equation (1) contains four independent parameters (Θ_s, Θ_r, α, N) that have to be estimated (h is assumed to be positive). Equation (2)
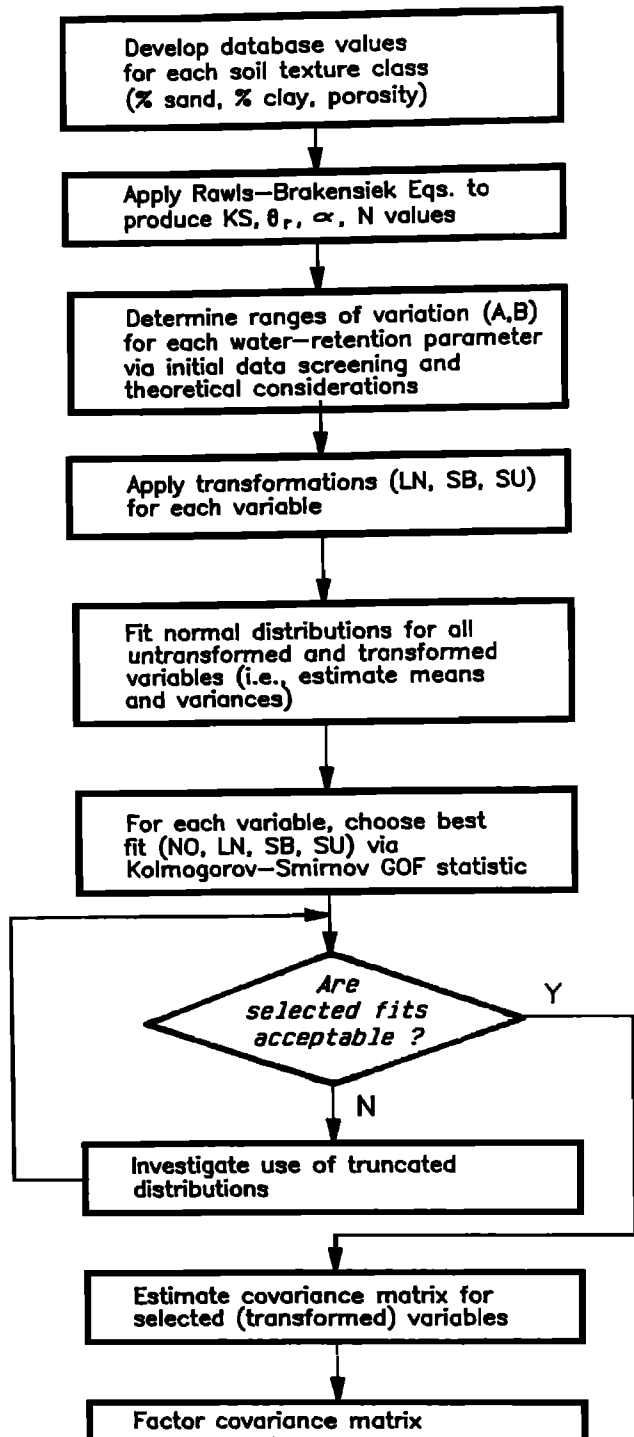

Fig. 3. Procedure used for identification, fitting, and estimation of parameter distributions.

TABLE 1. Descriptive Statistics for Percent Sand and Clay Content

Soil Type	Sand				Clay			
	$\bar{x}$	s	CV	n	$\bar{x}$	s	CV	n
Clay*	14.9	10.7	71.6	1177	55.2	10.9	19.7	1177
Clay loam	29.8	5.9	19.7	1317	32.6	3.7	11.4	1317
Loam	40.0	6.5	16.3	1991	19.7	5.2	26.3	1991
Loamy sand	80.9	3.8	4.6	881	6.4	3.2	50.1	881
Silt	5.8	4.5	77.2	115	9.5	2.7	28.9	115
Silt loam	16.6	11.7	70.8	3050	18.5	5.9	31.6	3050
Silty clay	6.1	4.5	73.5	1002	46.3	4.9	10.7	1002
Silty clay loam	7.6	5.3	70.7	1882	33.2	3.7	11.1	1882
Sand	92.7	3.7	4.0	803	2.9	2.0	67.1	803
Sandy clay	47.5	3.9	8.2	74	41.0	4.5	10.9	74
Sandy clay loam	54.3	7.3	13.5	610	27.4	4.0	14.6	610
Sandy loam	63.4	7.9	12.5	2835	11.1	4.8	43.2	2835

Here, $\bar{x}$, mean; s, standard deviation; CV, coefficient of variation (percent); and n, sample size.
*Agricultural soil, less than 60% clay.

contains one additional parameter, K_s, that has to be estimated.

A soil database compiled by *Carsel et al.* [1988] was used to obtain bulk density, sand, and clay contents for the 12 Soil Conservation Service (SCS) textural classifications including: clay, clay loam, silt, silt loam, silty clay, silty clay loam, sand, sandy clay, sandy clay loam, and sandy loam. These data were obtained from measurements for all soils reported in SCS Soil Survey Information Reports. These reports (published by State) generally contain static soils data for the predominant soil series within a state. A total of 42 books representing 42 states were used to develop the database. Saturated water content was inferred from bulk density [*Rawls and Brakensiek*, 1985]. The saturated water contents, the sand contents, and the clay contents reported for each of the SCS classifications then were used to compute saturated hydraulic conductivity (centimeters per hour) and water retention parameters for the *van Genuchten* [1976] model using a multiple regression equation developed by *Rawls and Brakensiek* [1985]. The general form of the regression equation (where f denotes any of the variables ln (K_s), Θ_r, ln (α^{-1}), or ln $(N - 1)$) and related coefficients are provided in Figure 1. Their work included testing

of the regression model using 95 soils with textural classifications ranging from clays to sands. Estimated means for final infiltration rates of each soil were within one standard deviation of the observed means. The regression equations were developed for natural soils only; modifications would be necessary for soils having temporal variations such as surface crusts, etc. Spatially, the hydraulic parameters are expected to vary with the percent sand, clay, and saturated water content. By applying the equations to each SCS soil classification with large deviations of percent sand, clay, and saturated water contents, spatial representation of hydraulic parameters can be estimated.

Statistical Analysis Procedures

The database of computed saturated hydraulic conductivities (K_s) and *van Genuchten* [1976] water retention parameters (Θ_r, α, N) for each of the 12 soil textural classifications was used as the basis for characterization of probability distributions for these variables. Descriptive statistics, moments, and other distributional characteristics were examined. Empirical cumulative distribution functions (CDF) were derived

TABLE 2. Descriptive Statistics for Saturated Water Content θ_s

Soil Type	Saturated Water Content θ_s			
	$\bar{x}$	s	CV	n
Clay*	0.38	0.09	24.1	400
Clay loam	0.41	0.09	22.4	364
Loam	0.43	0.10	22.1	735
Loamy sand	0.41	0.09	21.6	315
Silt	0.46	0.11	17.4	82
Silt loam	0.45	0.08	18.7	1093
Silty clay	0.36	0.07	19.6	374
Silty clay loam	0.43	0.07	17.2	641
Sand	0.43	0.06	15.1	246
Sandy clay	0.38	0.05	13.7	46
Sandy clay loam	0.39	0.07	17.5	214
Sandy loam	0.41	0.09	21.0	1183

Here, $\bar{x}$, mean; s, standard deviation; CV, coefficient of variation (percent); and n, sample size.
*Agricultural soil, less than 60% clay.

TABLE 3. Descriptive Statistics for Residual Water Content θ_r

Soil Type	Residual Water Content θ_r			
	$\bar{x}$	s	CV	n
Clay*	0.068	0.034	49.9	353
Clay loam	0.095	0.010	10.1	363
Loam	0.078	0.013	16.5	735
Loamy sand	0.057	0.015	25.7	315
Silt	0.034	0.010	29.8	82
Silt loam	0.067	0.015	21.6	1093
Silty clay	0.070	0.023	33.5	371
Silty clay loam	0.089	0.009	10.9	641
Sand	0.045	0.010	22.3	246
Sandy clay	0.100	0.013	12.9	46
Sandy clay loam	0.100	0.006	6.0	214
Sandy loam	0.065	0.017	26.6	1183

Here, $\bar{x}$, mean; s, standrad deviation; CV, coefficient of variation (percent); and n, sample size.
*Agricultural soil, less than 60%.

TABLE 4. Descriptive Statistics for Hydraulic Conductivity K_S and *van Genuchten* [1976] Water Retention Parameter α

Soil Type	Hydraulic Conductivity K_S, cm hr^{-1}				α, cm^{-1}			
	$\bar{x}$	s	CV	n	$\bar{x}$	s	CV	n
Clay*	0.20	0.42	210.3	114	0.008	0.012	160.3	400
Clay loam	0.26	0.70	267.2	345	0.019	0.015	77.9	363
Loam	1.04	1.82	174.6	735	0.036	0.021	57.1	735
Loamy sand	14.59	11.36	77.9	315	0.124	0.043	35.2	315
Silt	0.25	0.33	129.9	88	0.016	0.007	45.0	82
Silt loam	0.45	1.23	275.1	1093	0.020	0.012	64.7	1093
Silty clay	0.02	0.11	453.3	126	0.005	0.005	113.6	126
Silty clay loam	0.07	0.19	288.7	592	0.010	0.006	61.5	641
Sand	29.70	15.60	52.4	246	0.145	0.029	20.3	246
Sandy clay	0.12	0.28	234.1	46	0.027	0.017	61.7	46
Sandy clay loam	1.31	2.74	208.6	214	0.059	0.038	64.6	214
Sandy loam	4.42	5.63	127.0	1183	0.075	0.037	49.4	1183

Here, $\bar{x}$, mean; s, standard deviation; CV, coefficient of variation (percent); and n, sample size. *Agricultural soil, less than 60%.

for all of these variables, and hypothesized distributions were fitted. The first objective was to obtain the set of best fitting distributions that would adequately approximate the empirical distributions. In each instance, a mathematical transformation was sought that would produce a normally distributed variable.

Fitted CDF were selected generally from a class of transformed normal distributions known as the Johnson system [*Johnson and Kotz*, 1970]. The normal (Gaussian) distribution (denoted by NO) also was used in those cases where no transformation was necessary to achieve normality. The Johnson system involves three main distribution types: LN, lognormal; SB, log ratio; and SU, hyperbolic arcsine. By definition, a random variable that has a lognormal distribution will have a normal distribution after applying a logarithmic transformation. Similarly, variables following the SB or SU distributions also can be transformed to normality, as is described below. Oftentimes, the lognormal distribution is inadequate for the representation of given empirical data, whereas the SB or SU may be well-suited. All of the types of Johnson distributions represent transformations of variables that have normal distributions after the transformations are applied. From an empirical standpoint, one would choose the transformation that does the best job of producing normally distributed data in any given case. The underlying reason why one transformation might work better than another in this regard actually is related to certain other characteristics of the distribution.

The third standard moment (termed skewness, denoted by $(\beta_1)^{1/2}$) and fourth standard moment (termed kurtosis, denoted by β_2) of Johnson family distributions can be used to discriminate among the three types. Geometrically, the plane defined by the set of all values of β_1 and β_2 divides into two regions: one corresponding to SB distributions, the other corresponding to SU distributions (see Figure 2). The skewness and kurtosis of the lognormal are functions of the variance (σ^2) of the log transform of the random variable (i.e., σ^2 denotes the variance of the normal distribution that obtains for log x when X is lognormally distributed). Hence the boundary defined by the parametric equations

$$\beta_1 = (w - 1)(w + 2)^2 \qquad \beta_2 = w^4 + 2w^3 + 3w^2 - 3 \qquad (3)$$

where $w = \exp(\sigma^2)$ is the locus of all points in the plane that correspond to lognormal distributions. For given skewness values, the region where kurtosis is less than that of the lognormal is the SB region; the region with greater kurtosis is the SU region. The SB region is bounded also by the line

$$\beta_2 = \beta_1 + 1$$

the limit for all distributions. Each point in the (β_1, β_2) plane is uniquely associated with a specific Johnson distribution. For empirical data, there often will be an SB or an SU distribution that fits better than the lognormal, a situation encountered when the (β_1, β_2) point lies far from the lognormal boundary line. As the skewness and kurtosis coefficients approach 0 and 3, respectively, the limiting distribution is the normal.

The Johnson transformations may be given as

$$LN: \quad Y = \ln(X) \qquad (4)$$

$$SB: \quad Y = \ln[(X - A)/(B - X)] \qquad (5)$$

$$SU: \quad Y = \sinh^{-1}[U] = \ln[U + (1 + U^2)^{1/2}] \qquad (6)$$

TABLE 5. Descriptive Statistics for *van Genuchten* [1976] Water Retention Model Parameter N

Soil Type	N			
	$\bar{x}$	s	CV	n
Clay*	1.09	0.09	7.9	400
Clay loam	1.31	0.09	7.2	364
Loam	1.56	0.11	7.3	735
Loamy sand	2.28	0.27	12.0	315
Silt	1.37	0.05	3.3	82
Silt loam	1.41	0.12	8.5	1093
Silty clay	1.09	0.06	5.0	374
Silty clay loam	1.23	0.06	5.0	641
Sand	2.68	0.29	20.3	246
Sandy clay	1.23	0.10	7.9	46
Sandy clay loam	1.48	0.13	8.7	214
Sandy loam	1.89	0.17	9.2	1183

Here, $\bar{x}$, mean; s, standard deviation; CV, coefficient of variation (percent); and n, sample size. *Agricultural soil, less than 60%.

TABLE 6. Statistical Parameters Used for Distribution Approximation

Soil Texture	Hydraulic Variable	Limits of Variation A	Limits of Variation B	Trans-formation	Estimated* Mean	Estimated* Standard Deviation	D†	Truncation Limits on Transformed Variable	
S	K_S	0.	70.	SB	− 0.394	1.15	0.045		
S	θ_r	0.	0.1	LN	− 3.12	0.224	0.053		
S	α	0.	0.25	SB	0.378	0.439	0.050		
S	N	1.5	4.0	LN	0.978	‾0.100	0.063		
SL	K_S	0.	30.	SB	2.49	1.53	0.029		
SL	θ_r	0.00	0.11	SB	0.384	0.700	0.034		
SL	α	0.00	0.25	SB	− 0.937	0.764	0.044		
SL	N	1.35	3.00	LN	0.634	0.082	0.039		
LS	K_S	0.	51.	SB	− 1.27	1.40	0.036		
LS	θ_r	0.	0.11	SB	0.075	0.567	0.043		
LS	α	0.	0.25	NO	0.124	0.043	0.027		
LS	N	1.35	5.00	SB	− 1.11	0.307	0.070		
SIL	K_S	0.	15.	LN	− 2.19	1.49	0.046		
SIL	θ_r	0.00	0.11	SB	0.478	0.582	0.073		
SIL	α	0.00	0.15	LN	− 4.10	0.555	0.083		
SIL	N	1.	2.	SB	− 0.370	0.526	0.104		
SI	K_S	0.	2.	LN‡	− 2.20	0.700	0.168	− 2.564	− 0.337
SI	θ_r	0.0	0.09	ND‡	0.042	0.015	0.089	0.013	0.049
SI	α	0.0	0.1	NO	0.017	0.006	0.252		
SI	N	1.2	1.6	NO	1.38	0.037	0.184		
C	K_S	0.	5.	SB	− 5.75	2.33	0.122		
C	θ_r	0.0	0.15	SU‡	0.445	0.282	0.058	.0065	0.834
C	α	0.0	0.15	SB‡	− 4.145	1.293	0.189	− 5.01	0.912
C	N	0.9	1.4	LN‡	0.0002	0.118	0.131	0.	0.315
SIC	K_S	0.	1.	LN	− 5.69	1.31	0.205		
SIC	θ_r	0.00	0.14	NO	0.070	0.023	0.058		
SIC		0.00	0.15	LN	− 5.66	0.584	0.164		
SIC	N	1.0	1.4	SB	− 1.28	0.821	0.069		
SC	K_S	0.0	1.5	LN	− 4.04	2.02	0.130		
SC	θ_r	0.00	0.12	SB	1.72	0.700	0.078		
SC		0.00	0.15	LN	− 3.77	0.563	0.127		
SC	N	1.0	1.5	LN	0.202	0.078	0.100		
SICL	K_S	0.0	3.5	SB	− 5.31	1.62	0.049		
SICL	θ_r	0.0	0.115	NO	0.088	0.009	0.056		
SICL	α	0.0	0.15	SB	− 2.75	0.605	0.082		
SICL	N	1.0	1.5	NO	1.23	0.061	0.082		
CL	K_S	0.	7.5	SB‡	− 5.87	2.92	0.058	− 8.92	2
CL	θ_r	0.	0.13	SU	0.679	0.060	0.061		
CL	α	0.	0.15	LN	− 4.22	0.72	0.052		
CL	N	1.0	1.6	SB	0.132	0.725	0.035		
SCL	K_S	0.	20.	SB	− 4.04	1.85	0.047		
SCL	θ_r	0.00	0.12	SB‡	1.65	0.439	0.077	0.928	2.94
SCL	α	0.00	0.25	SB	− 1.38	0.823	0.048		
SCL	N	1.	2.	LN	0.388	0.086	0.043		
L	K_S	0.	15.	SB	− 3.71	1.78	0.019		
L	θ_r	0.	0.12	SB	0.639	0.487	0.064		
L	α	0.	0.15	SB	− 1.27	0.786	0.039		
L	N	1.	2.	SU	0.532	0.099	0.036		

S, sand; SL, sandy loam; LS, loamy sand; SIL, silt loam; SI, silt; C, clay; SIC, silty clay; SC, sandy clay; SICL, silty clay loam; CL, clay loam; SCL, sandy clay loam; and L, loam.
*For distribution of transformed variables.
†Kolmogorov-Smirnov goodness-of-fit test statistic.
‡Truncated form of the distribution.

where ln denotes natural log, X denotes an untransformed variable with limits of variation from A to B $(A < X < B)$, and $U = (X − A)/(B − A)$. This form of the LN distribution is defined for all positive values of X, being unbounded above. SB is bounded between limits A and B, while SU generally is unbounded. The use of A and B in the SU transformation is for mathematical convenience. In the present application, X corresponds to any of the variables K_S, Θ_r, α, N. In each case, Y has a normal distribution.

In using this approach, the limits of variation (A and B) for each variable (K_S, Θ_r, α, N) were determined a priori on the basis of observed data ranges and theoretical considerations and, then, utilized in the three LN-SB-SU transformations (equations (4), (5), and (6)). Generally, the third and fourth sample moments (skewness and kurtosis) can be used to determine which of the three distribution types of the Johnson family is an appropriate choice in any given case, as is noted above. In this application, however, where only empirical fits were needed, it was sufficient and convenient to independently fit the normal distribution to the original data set and to the

TABLE 7. Correlations Among Transformed Variables
Presented With the Factored Covariance Matrix

	K_S	θ_r	α	N
		Silt (n = 61)		
K_S	0.535	− 0.002	0.003	0.013
θ_r	− 0.204	0.008	0.000	− 0.015
α	0.984	− 0.200	0.001	0.014
N	0.466	− 0.610	0.551	0.013
		Clay (n = 95)		
K_S	1.96	0.070	0.565	0.048
θ_r	0.972	0.017	− 0.080	− 0.014
α	0.948	0.890	0.172	0.002
N	0.908	0.819	0.910	0.016
		Silty Clay (n = 123)		
K_S	1.25	0.008	0.314	0.367
θ_r	0.949	0.003	0.040	− 0.086
α	0.974	0.964	0.060	0.066
N	0.908	0.794	0.889	0.131
		Sandy Clay (n = 46)		
K_S	2.02	0.883	0.539	0.076
θ_r	0.939	0.324	0.063	0.004
α	0.957	0.937	0.150	− 0.001
N	0.972	0.928	0.932	0.018
		Sand (n = 237)		
K_S	1.04	− 0.109	0.328	0.081
θ_r	− 0.515	0.182	0.258	− 0.047
α	0.743	0.119	0.143	− 0.011
N	0.843	− 0.858	0.298	0.017
		Sandy Loam (n = 1145)		
K_S	1.60	− 0.153	0.037	0.211
θ_r	− 0.273	0.538	0.017	− 0.194
α	0.856	0.151	0.014	0.019
N	0.686	− 0.796	0.354	0.108
		Loamy Sand (n = 313)		
K_S	1.48	− 0.201	0.037	0.211
θ_r	− 0.359	0.522	0.017	− 0.194
α	0.986	− 0.301	0.014	0.019
N	0.730	− 0.590	0.354	0.108
		Silt Loam (n = 1072)		
K_S	1.478	− 0.201	0.525	0.353
θ_r	− 0.359	0.522	0.030	− 0.170
α	0.986	− 0.301	0.082	0.234
N	0.730	− 0.590	0.775	0.158
		Silty Clay Loam (n = 591)		
K_S	1.612	0.006	0.511	0.049
θ_r	0.724	0.005	0.048	− 0.009
α	0.986	0.77	0.073	0.008
N	0.918	0.549	0.911	0.017
		Clay Loam (N = 328)		
K_S	1.92	0.040	0.589	0.542
θ_r	0.790	0.031	− 0.062	− 0.154
α	0.979	0.836	0.106	0.065
N	0.936	0.577	0.909	0.116
		Sandy Clay Loam (n = 212)		
K_S	1.85	0.102	0.784	0.077
θ_r	0.261	0.378	0.122	− 0.031
α	0.952	0.392	0.220	− 0.008
N	0.909	− 0.113	0.787	0.016
		Loam (n = 664)		
K_S	1.41	− 0.100	0.611	0.055
θ_r	0.204	0.478	0.073	− 0.055
α	0.982	− 0.086	0.093	0.026
N	0.632	− 0.748	0.591	0.029

Entries in the lower triangular portion of the matrix are sample Pearson product-moment correlations. The diagonal and upper triangular entries form the triangular Cholesky decomposition of the sample covariance matrix. N, sample size.

three sets generated by applying the Johnson transformations on a given variable. In each case, the first two moments (i.e., mean and variance) of the transformed values were used to estimate the corresponding parameters of a normal distribution. An objective measure of goodness of fit, the Kolmogorov-Smirnov (K-S) D statistic, then was used to select the best fitting distribution from among the four candidates (NO, LN, SB, SU). The K-S D statistic is defined basically as the maximum observed deviation between an empirical CDF and a fitted CDF, so that the smallest observed value of D signified the most appropriate transformation in any given case.

In following this procedure, a fitted normal or Johnson distribution was derived for each variable within each soil textural class. A knowledge of the type of transformation, the estimated mean and variance of the associated normal distribution, and the limits of variation were sufficient to define completely the fitted distribution of any given variable.

Data peculiarities, such as outliers, required that the fits be carefully scrutinized with respect to proper estimation of parameters. For the most part, maximum likelihood estimates of the mean and variance were computed on the basis of complete data sets, although in a few cases trimmed estimates were utilized. The goodness-of-fit criterion was based consistently on untrimmed data sets to ensure objectivity. That is, in some instances, outlying values were not used in estimating the mean and variance, but they were included for goodness-of-fit calculations.

If a case exhibited characteristics that usually are associated with truncated distributions to such a degree that a nontruncated fit was considered unacceptable, efforts were made to use truncated Johnson system distributions. In these situations, maximum likelihood estimates of the mean and variance of the normal parent distribution for the transformed variable were obtained using methods appropriate for the doubly truncated normal [*Johnson and Kotz*, 1970]. That is, a truncated normal distribution was fitted to the transformed data in these cases. The density function of a truncated normal distribution can be given as

$$f_T(x) = f(x)/[F(b) - F(a)]$$
$$= \sigma^{-1}\phi\,[(x - \mu)/\sigma]\Phi_{b-a}^{-1} \qquad a \le x \le b \qquad (7)$$

where

$$f(x) = \sigma^{-1}(2\pi)^{-1/2} \exp\left\{-(1/2)[(x - \mu)/\sigma]^2\right\} \qquad (8)$$

is the density function of the nontruncated normal parent distribution having mean μ, and variance σ^2 and associated CDF $F(x)$, where

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2) \qquad (9)$$

is the standard normal density with associated CDF $\Phi(z)$ and where

$$\Phi_{b-a}^{-1} = \Phi[(b - \mu)/\sigma] - \Phi[(a - \mu)/\sigma] \qquad (10)$$

The limits of truncation are a and b. The first two moments of the truncated normal distribution are related mathematically to the moments of the nontruncated parent distribution through two nonlinear equations. The expressions for the expected value and variance of the truncated distribution are

$$\mu_T = \mu + \sigma(\phi_a - \phi_b)\Phi_{b-a}^{-1} \qquad (11)$$

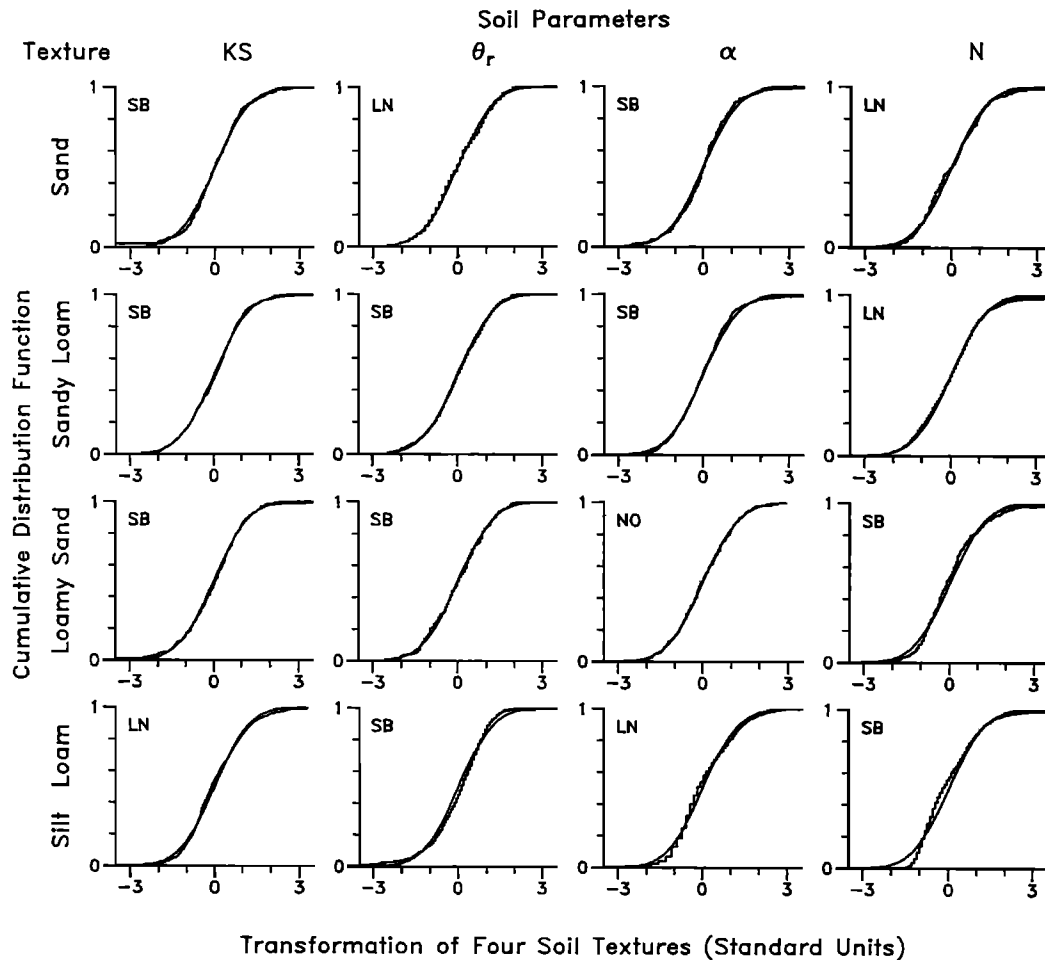$$\sigma_T^2 = \sigma^2\{1 + (z_a\phi_a - z_b\phi_b)\Phi_{b-a}^{-1} - (\phi_a - \phi_b)\Phi_{b-a}^{-2}\} \qquad (12)$$

Fig. 4. Observed and predicted cumulative distributions for saturated hydraulic conductivity K_S and van Genuchten [1976] model parameters Θ_r, α, and N for sand, sandy loam, loamy sand, and silty loam soils. NO, normal; LN, lognormal; SU, hyperbolic arcsine; SB, log ratio.

where $z_a = (a - \mu)/\sigma$, $z_b = (b - \mu)/\sigma$, $\phi_a = \phi(z_a)$, and $\phi_b = \phi(z_b)$. By equating these nonlinear expressions to sample moments obtained from the data and, then, solving numerically, estimates of the parameters μ and σ of the parent distribution were determined. This estimation procedure was used only in those cases where truncated normal distributions were needed.

After choosing the best fitting distribution, sample covariances (and correlations) among the selected transformed variables were computed. These served to estimate the covariances needed by a joint multivariate distribution model. Since the Johnson system provides a mechanism for developing, after transformations, a set of normally distributed variables, a multivariate normal distribution model was selected to represent the joint probability density for the transformed variables. In this manner, the estimated covariance structure was incorporated for future use in Monte Carlo simulations. (A one-dimensional finite element solute transport model with a Monte Carlo preprocessor is currently being developed.) This identification fitting estimation procedure is summarized in Figure 3. It depicts the steps of determining appropriate transformations and corresponding estimates of distribution means, variances, and covariances.

The multivariate normal distribution is parameterized in terms of marginal distribution means and variances and pair-

wise covariances in the form of a covariance matrix. The multivariate normal density function is given by

$$f(z; \mu, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \{-(1/2)(z - \mu)'\Sigma^{-1}(z - \mu)\}$$

$$-\infty < z_i < \infty \qquad i = 1, 2, \cdots, p \qquad (13)$$

where z represents a vector of p random variables with mean vector μ and variance-covariance matrix Σ, and $|\Sigma|$ denotes the determinant of Σ. Random deviates from a correlated multivariate normal distribution can be produced by first generating a vector z of independent standard normal deviates and then applying a linear transformation of the form

$$y = \mu + T'z \qquad (14)$$

where μ is the desired vector of means and T' is the transpose of an upper triangular matrix derived from the factored form of the symmetric covariance matrix $\Sigma = T'T$. The existence of this factorization requires that the covariance matrix be positive definite.

Within each soil textural class, after transformations were selected and distributions were fitted for all variables, sample Pearson product-moment correlations and covariances were calculated for the transformed variables, as is described above. These estimates were based on sets of complete observations
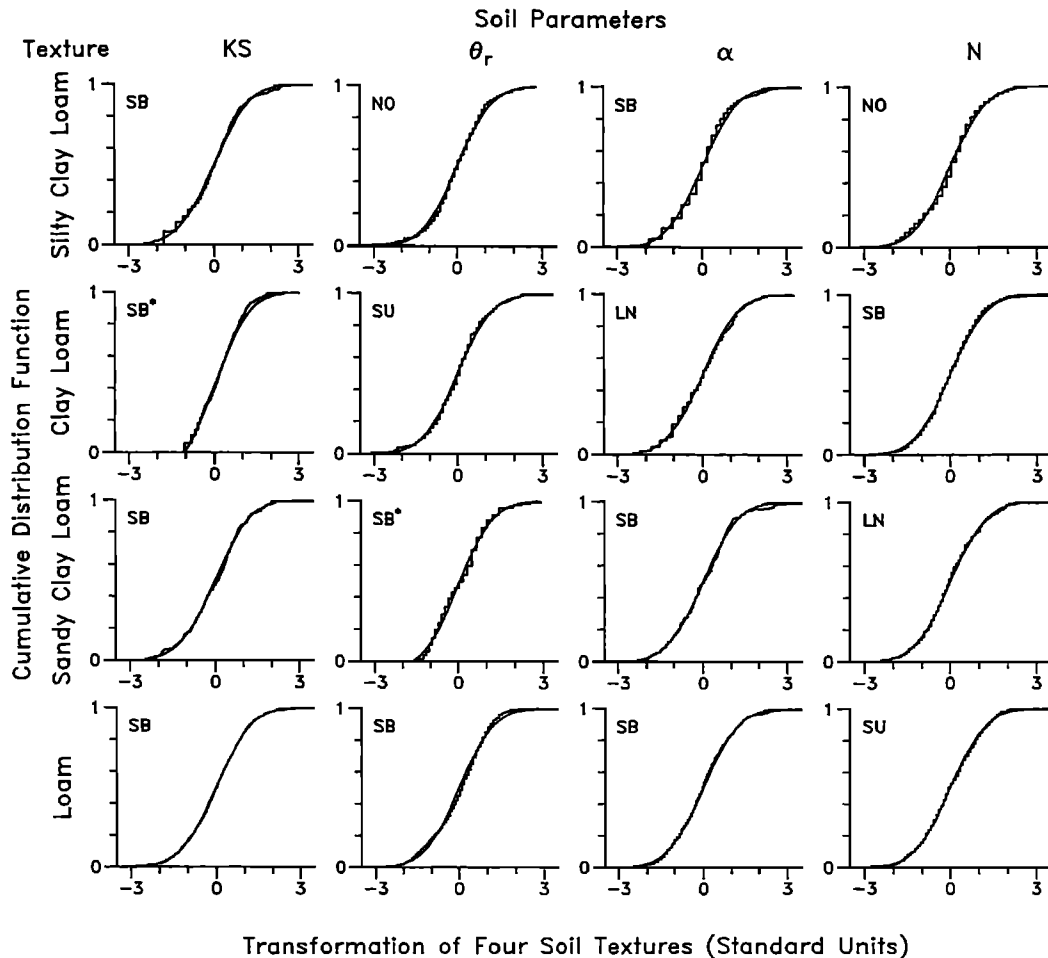
Fig. 5.   Observed and predicted cumulative distributions for saturated hydraulic conductivity K_S and van Genuchten [1976] model parameters Θ_r, α, and N for silty clay loam, clay loam, sandy clay loam, and loam soils. *NO*, normal; *LN*, lognormal; *SU*, hyperbolic arcsine; *SB*, log ratio. Asterisks indicate the truncated form.

in which all variables had nonmissing values. This approach utilized somewhat less information when compared to individual pairwise estimates, but it guaranteed that the covariance matrices would be positive definite and thus could be factored. The Cholesky decomposition algorithm [*Kennedy and Gentle*, 1980] was used to factor the estimated covariance matrices.

RESULTS AND DISCUSSION

Descriptive statistics for saturated water, sand, and clay contents, are provided in Tables 1 and 2. Estimated saturated hydraulic conductivity K_S and van Genuchten [1976] water retention parameters (Θ_r, α, N) are provided in Tables 3–5.

The CV for saturated water content Θ_s was less than 25% for all soil types. These values are consistent with those reported elsewhere [e.g., *Jury*, 1985]; therefore variability for saturated water content is minimal. The CV for percent sand was greater than 50% for clay, silt, silt loam, silty clay, and silty clay loam soils and less than 20% for clay loam, loam, loamy sand, sand, sandy clay, sand clay loam, and sandy loams. The CV for clay content was greater than 40% for loamy sand, sand, and sandy loam soils and less than 35% for clay, clay loam, loam, silt, silt loam, silty clay, silty clay loam, and sand clays. Generally, the CV for simulated values of residual water content Θ_r and the van Genuchten [1976]

model parameter N were less than 30 and 20%, respectively. Higher CV were observed for the van Genuchten [1976] model parameter value α. The CV was generally greater than 35% for α (CV for sand was 20.3). The CV for simulated saturated hydraulic conductivity K_S ranged from 453.3 for silty clays to 52.4 for sands. Common agricultural soils such as silt loams, loamy sands, loams, and sandy loams exhibited CV of 174.6, 77.9, 275.1, and 127.0, respectively. These values compare favorably to measured CV for loamy sands, sandy loams, sands, silty clays, and silty clay loams of 69–105, 178–190, 69, 92–320, and 48–118, respectively [*Smith et al.*, 1987]. Sensitivity of the characteristic curve (equation (1)) as indicated by the CV would appear to be generally related to α.

The results for hydraulic conductivity and van Genuchten [1976] model parameters indicated that considerable differences are expected for any simulation of solute movement in the unsaturated zone. It follows that uncertainty estimates should be incorporated into any associated modeling study.

Estimates of the distribution mean and standard deviation for appropriately transformed variables, limits of variation for the original variables, and values of the K-S goodness-of-fit statistic D (maximum absolute deviation between the empirical and fitted CDF) are displayed in Table 6. In cases where truncated distributions were used, the truncation limits also
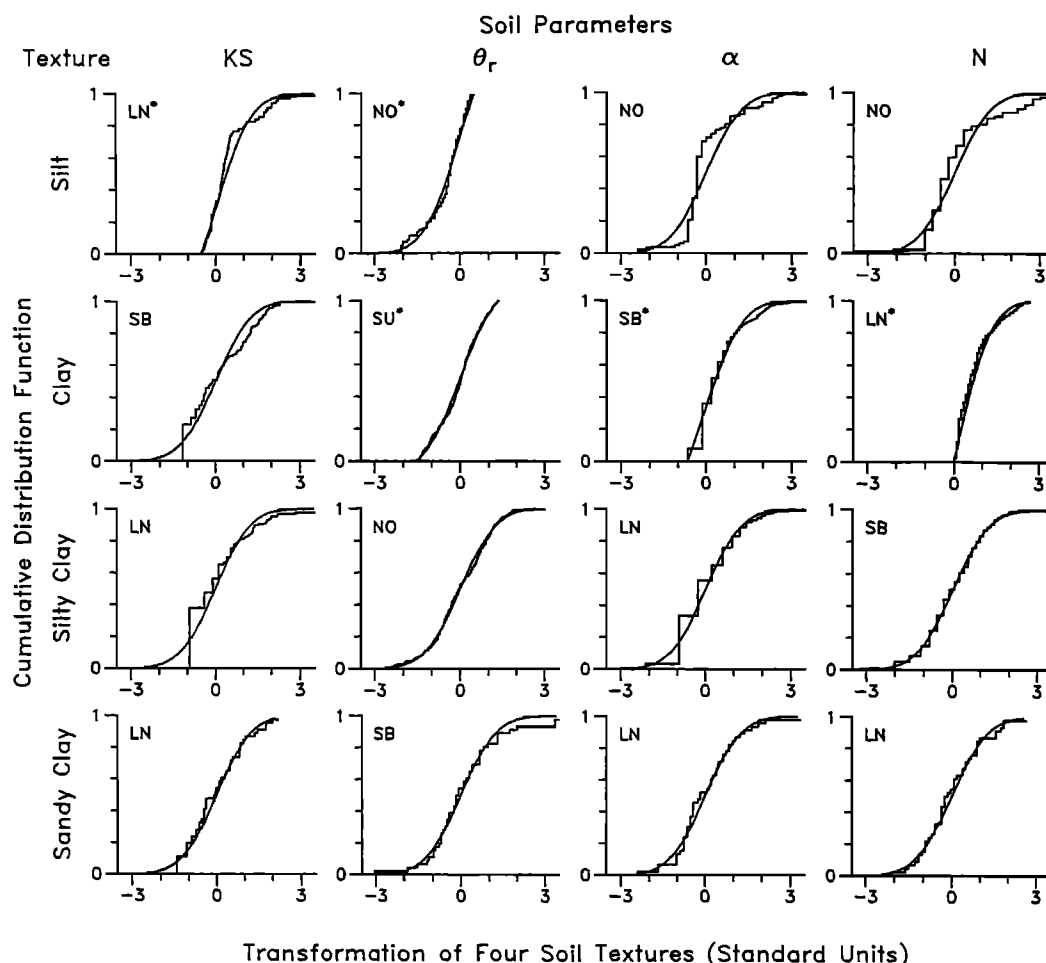
Fig. 6. Observed and predicted cumulative distributions for saturated hydraulic conductivity K_S and van Genuchten [1976] model parameters Θ_r, α, and N for silt, clay, silty clay, and sandy clay soils. NO, normal; LN, lognormal; SU, hyperbolic arcsine; SB, log ratio. Asterisks indicate the truncated form.

are shown. Correlations among transformed variables are given in Table 7; these appear (in boldface) as the six entries below the matrix diagonal in each case. The entries on and above the diagonal comprise the upper triangular matrix that forms a factor of the estimated covariance matrix.

Figures 4–6 display plots of the empirical and fitted CDF for transformed values of saturated hydraulic conductivity and van Genuchten [1976] model parameters for each of the 12 soil textural classes. In each case, standardized (zero mean and unit variance) scaling of the transformed variable is utilized for purposes of uniform presentation.

The fitted CDF are considered acceptable for simulation because they are based on available data and because the absolute errors between observed and predicted CDF are not expected to be distinguishable when using solute transport models. The present approach met the primary objective of obtaining good approximations for most of the underlying distributions. In addition, the fitted distributions had a smoothing effect in cases where data gaps may have occurred (e.g., silt, clay, and sandy clay soils). This offers an implicit advantage for data representation.

Very few of the data sets could be adequately described by the normal distribution without using one of the Johnson transformations. The SCS textural triangle system used for classifying soils is thought to have contributed to the existence of truncated forms and of forms having properties un-

characteristic of the normal distribution. Classification schemes and/or restrictions that weight results could have produced the results observed for these soils. Notably, many data sets were significantly better described by the SB and SU distributions rather than the more commonly used lognormal, although the lognormal was selected in about one third of the cases. Truncated distributions were used in 7 of the 48 cases; the resulting fits were significantly improved in each of these.

In most cases, correlations were significant for the van Genuchten [1976] model parameters. For example, correlations generally were greater than 0.70 for between K_S and α, and between K_S and N. The implication is that an assumption of independence in a Monte Carlo simulation is not plausible. (Such an assumption would add considerable white noise to the results, thus limiting their utility.)

The Monte Carlo implementation of these results would require that multivariate normal deviates be generated with means, variances, and covariances as previously estimated. These random values then would be inverse-transformed into the original scaling used for the hydraulic parameters. Equations (4)–(6) may be inverted mathematically to produce

$$LN: \quad X = \exp(Y) \tag{15}$$

$$SB: \quad X = [B \exp(Y) + A]/[1 + \exp(Y)] \tag{16}$$

$$SU: \quad X = A + (B - A)[\exp(Y) - \exp(-Y)]/2 \tag{17}$$

```
INPUT: Limits of variation (A, B)
       Transformations (NO, LN, SB, SU)
       Mean estimates (u)
       Factored covariance matrix (T)
```

ly generated vector would be retained only when all range constraints are satisfied.

Figure 7 illustrates Monte Carlo implementation using the multivariate normal–Johnson transformation approach. The data provided in Tables 6 and 7 can be used to select and parameterize the distributions for K_S, Θ_r, α, and N in any given soil textural class. The factored covariance matrices and transformed variable means can be used to generate sets of correlated normal random deviates. These values would then be translated mathematically, as in (15)–(17), depending on the fitted distributions, to produce random values for the soil water retention parameters.

Example

For the silt loam soil data, the limits of variation, variable transformations, and estimates of the transformed variable means are obtained from Table 6, and the factored covariance matrix is read from Table 7:

K_S: $y_1 = \ln (K_S)$ $A = 0, B = 15.00, LN$

Θ_r: $y_2 = \ln [\Theta_r/(0.11 - \Theta_r)]$ $A = 0, B = 0.11, SB$

α: $y_3 = \ln (\alpha)$ $A = 0, B = 0.15, LN$

N: $y_4 = \ln [(N - 1)/(2 - N)]$ $A = 1, B = 2.00, SB$

$$u = \begin{bmatrix} -2.187 \\ 0.478 \\ -4.099 \\ -0.370 \end{bmatrix},$$

$$T = \begin{bmatrix} 1.475 & -0.201 & 0.525 & 0.353 \\ 0 & 0.522 & 0.030 & -0.170 \\ 0 & 0 & 0.082 & 0.234 \\ 0 & 0 & 0 & 0.158 \end{bmatrix}$$

Recall that the transformations were selected so that data associated with each y_i were approximately normally distributed. The vector of means u consists of the estimated means for the distributions of the transformed variables y_1, y_2, y_3, and y_4. The estimated covariance matrix for the y_i may be calculated from T as

$$S = T'T = \begin{bmatrix} 2.176 & -0.296 & 0.774 & 0.521 \\ -0.296 & 0.313 & -0.090 & -0.160 \\ 0.774 & -0.090 & 0.283 & 0.199 \\ 0.521 & -0.160 & 0.199 & 0.233 \end{bmatrix}$$

relation matrix, as given in Table 7, can be computed as $R = DSD$, where D is a diagonal matrix with elements equal to the reciprocals of the square roots of the diagonal elements of S.

To illustrate how a random vector of values $(K_S, \Theta_r, \alpha, N)$ is produced, suppose that $z' = (-0.592, -0.009, 1.011, -1.649)$ is a vector of independent standard normal deviates. A new random vector y is derived by application of the transformation $y = u + T'z$:

$$y = \begin{bmatrix} -2.187 \\ 0.478 \\ -4.099 \\ -0.370 \end{bmatrix} + \begin{bmatrix} 1.475 & 0 & 0 & 0 \\ -0.201 & 0.522 & 0 & 0 \\ 0.525 & 0.030 & 0.082 & 0 \\ 0.353 & -0.170 & 0.234 & 0.158 \end{bmatrix}$$

$$\cdot \begin{bmatrix} -0.592 \\ -0.009 \\ 1.011 \\ -1.649 \end{bmatrix} = \begin{bmatrix} -3.060 \\ 0.592 \\ -4.327 \\ -0.602 \end{bmatrix}$$

Fig. 7. Procedure for implementing Monte Carlo numerical simulation using the multivariate normal Johnson transformation approach.

where Y represents a normally distributed variate with prescribed mean and variance. Of course, values of X generated in this manner generally must be checked to ensure that they are within the specified acceptable ranges (A to B) after transforming to original scales. In addition, whenever truncated normal distributions are involved, each element of the multivariate normal random deviate that is associated with a truncated distribution must be checked for range validity prior to inverse transformation. This is necessary, since the multivariate normal distribution model is parameterized in terms of the parent distributions of the truncated variables. A random-

```
'=================================================
' BASIC PROGRAM TO GENERATE RANDOM VALUES FOR SOIL PARAMETERS
'   (KS, QR, ALPHA, AND N) USING SILT LOAM INPUT DATA
'=================================================
'
DIM T(10),AMU(4),X(4),Y(4),Z(4),A(4),B(4),TR$(4),TA(4),TB(4)
'
'---------------------------------------------
'Load means, variable limits, transformations, and truncated
' limits, if any.  [Note: Code truncated distributions as
' "LN*", "SB*", "SU*", or "NO*" ]
'---------------------------------------------
'
DATA -2.187, 0.0, 15.0, "LN", 0., 0.
DATA 0.478, 0.0, 0.11, "SB", 0., 0.
DATA -4.099, 0.0, 0.15, "LN", 0., 0.
DATA -0.370, 1.0, 2.00, "SB", 0., 0.
'
FOR I=1 TO 4
  READ AMU(I), A(I), B(I), TR$(I), TA(I), TB(I)
  NEXT I
'
'---------------------------------------------
'Load factored covariance matrix T
'---------------------------------------------
'
DATA 1.4754, -0.2006, 0.5245, 0.3526
DATA        0.5215, 0.0300, -0.1696
DATA                0.0820, 0.2342
DATA                        0.1584
'
FOR I=1 TO 10
  READ T(I)
  NEXT I
'
'---------------------------------------------
'Get number to generate and open output file
'---------------------------------------------
'
INPUT "Enter number of vectors to generate . . . ", N
INPUT "Enter random number seed . . . . . . . . ", ISEED
'
RANDOMIZE ISEED
'
OPEN "MCARLO.SIL" FOR OUTPUT AS 1
'
'---------------------------------------------
'Begin loop
'---------------------------------------------
'
FOR L=1 TO N
'
'---------------------------------------------
'Generate independent normal random deviates
'---------------------------------------------
'
100 FOR J=1 TO 4
    Z(J) = -6.0
    FOR K=1 TO 12
    Z(J) = Z(J) + RND      'RND = uniform (0,1) deviate
    NEXT K
    NEXT J
```

```
'---------------------------------------------
'Apply linear transformation to produce correlated values
'---------------------------------------------
'
Y(1) = AMU(1) + T(1)*Z(1)
Y(2) = AMU(2) + T(2)*Z(1) + T(5)*Z(2)
Y(3) = AMU(3) + T(3)*Z(1) + T(6)*Z(2) + T(8)*Z(3)
Y(4) = AMU(4) + T(4)*Z(1) + T(7)*Z(2) + T(9)*Z(3) + T(10)*Z(4)
'
'---------------------------------------------
'Check limits for any truncated distributions
'---------------------------------------------
'
IF MID$(TR$(1),3,1)="*" THEN _
  IF Y(1)<TA(1) OR Y(1)>TB(1) THEN 100
IF MID$(TR$(2),3,1)="*" THEN _
  IF Y(2)<TA(2) OR Y(2)>TB(2) THEN 100
IF MID$(TR$(3),3,1)="*" THEN _
  IF Y(3)<TA(3) OR Y(3)>TB(3) THEN 100
IF MID$(TR$(4),3,1)="*" THEN _
  IF Y(4)<TA(4) OR Y(4)>TB(4) THEN 100
'
'---------------------------------------------
'Inverse transform correlated normals to get random deviates
' for KS, QR, ALPHA, N
'---------------------------------------------
'
FOR J=1 TO 4
  U = EXP(Y(J))
  IF MID$(TR$(J),1,2) = "LN" THEN _
    X(J) = U ELSE _
  IF MID$(TR$(J),1,2) = "SB" THEN _
    X(J) = (B(J)*U+A(J))/(1.0+U) ELSE _
  IF MID$(TR$(J),1,2) = "SU" THEN _
    X(J) = A(J) + 0.5*(B(J)-A(J))*(U-1.0/U) ELSE _
    X(J) = Y(J)
  NEXT J
'
'---------------------------------------------
'Ensure that values are within defined limits
'---------------------------------------------
'
IF X(1)<A(1) OR X(1)>B(1) THEN 100
IF X(2)<A(2) OR X(2)>B(2) THEN 100
IF X(3)<A(3) OR X(3)>B(3) THEN 100
IF X(4)<A(4) OR X(4)>B(4) THEN 100
'
'---------------------------------------------
'Output random vector (KS, QR, ALPHA, N) and close loop
'---------------------------------------------
'
PRINT #1, X(1); X(2); X(3); X(4)
'
NEXT L
'
'---------------------------------------------
'Finish
'---------------------------------------------
'
CLOSE #1
END
'
'---------------------------------------------
```

Fig. 8. BASIC program to generate random values for soil parameters using silt loam input data.

That is, y is a random vector from a multivariate normal distribution with mean u and variance-covariance matrix S. Inverse transformations (equations (15)–(17)) must be applied to y, as follows, in order to obtain the final random values for the original variables:

$$K_S = \exp(y_1) = \exp(-3.060) = 0.047$$

$$\Theta_r = [B \exp(y_2) + A]/[1 + \exp(y_2)]$$

$$= [0.11 \exp(0.592)]/[1 + \exp(0.592)] = 0.071$$

$$\alpha = \exp(y_3) = \exp(-4.327) = 0.013$$

$$N = [B \exp(y_4) + A]/[1 + \exp(y_4)]$$

$$= [2 \exp(-0.602) + 1]/[1 + \exp(-0.602)] = 1.354$$

These steps have been programmed as illustrated in Figure 8. By substituting appropriate values in the input data statements, any of the 12 soil textural classes may be represented. For the case of silt loam, the program was used to generate 1000 sets of values for K_S, Θ_r, α, and N. These data then were examined for agreement with the original observed data. These have been plotted in the form of double-bar histograms (Figures 9–12) showing both sets of relative frequencies for each variable. Table 8 displays the computed percentiles of both the generated values and the observed data for comparison.

CONCLUSIONS

A method was presented for developing probability density functions for several water retention characteristics for 12 soil
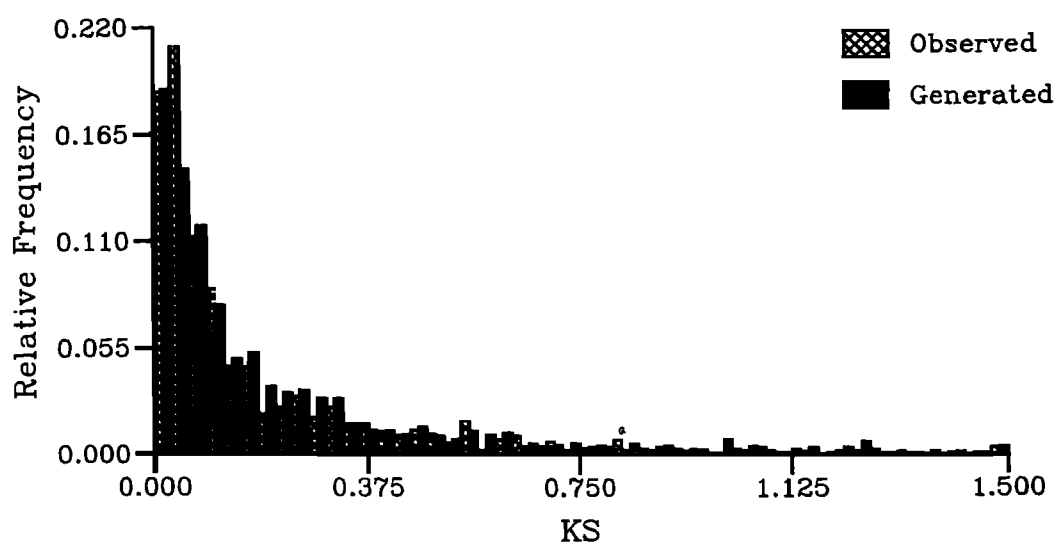
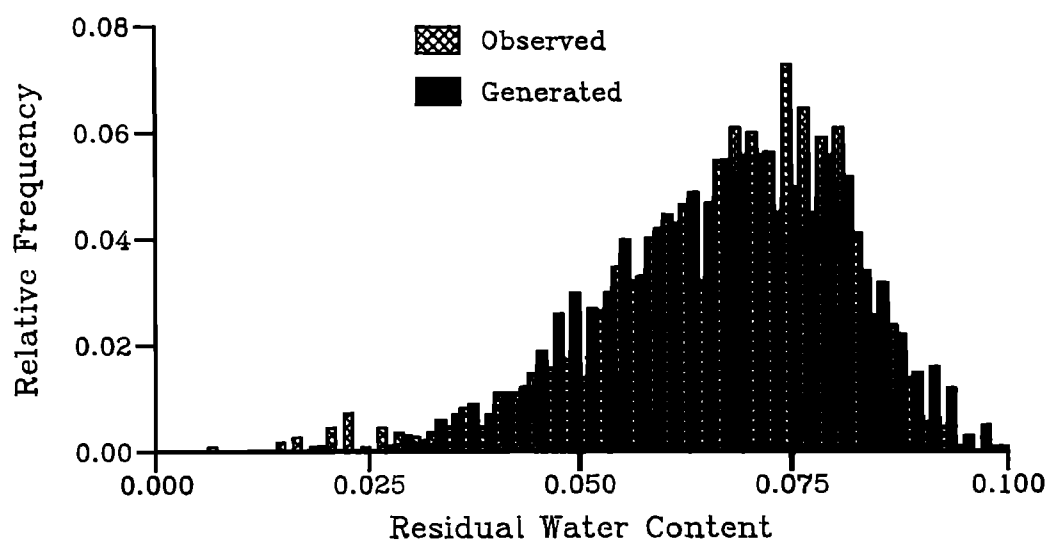Fig. 9. Histogram of randomly generated saturated hydraulic conductivity K_S values and original observed data.

Fig. 10. Histogram of randomly generated residual water content Θ_r values and original observed data.
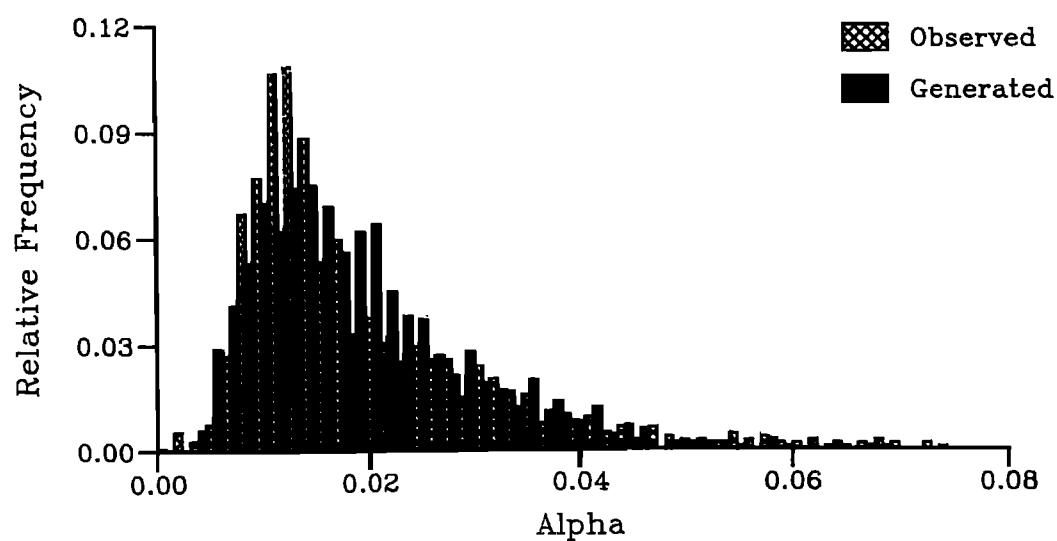

Fig. 11. Histogram of randomly generated *van Genuchten* [1976] water retention model parameter α values and original observed data.
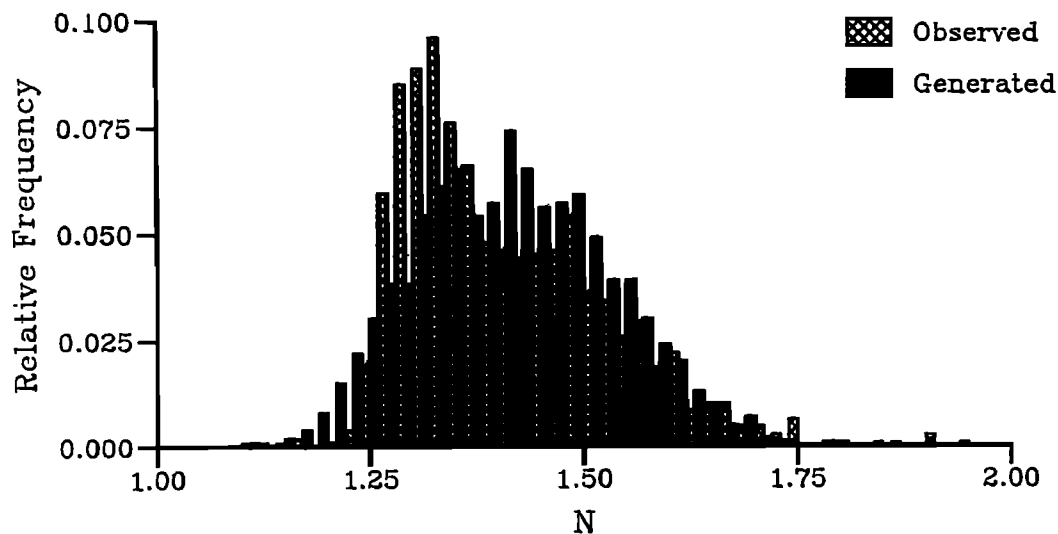
Fig. 12. Histogram of randomly generated van Genuchten [1976] water retention model parameter N values and original observed data.

TABLE 8. Percentiles of Generated and Observed Data, Silt Loam

Percentile Level	K_S		θ_r		α		N	
	Generated	Observed	Generated	Observed	Generated	Observed	Generated	Observed
1%	0.004	0.001	0.003	0.002	0.005	0.004	1.188	1.24
5%	0.009	0.011	0.043	0.040	0.007	0.008	1.238	1.27
10%	0.016	0.019	0.048	0.048	0.008	0.009	1.268	1.28
25%	0.044	0.040	0.057	0.059	0.012	0.011	1.331	1.32
50%	0.115	0.096	0.068	0.070	0.017	0.015	1.412	1.38
75%	0.316	0.310	0.078	0.078	0.024	0.025	1.495	1.49
90%	0.778	0.818	0.084	0.083	0.034	0.036	1.570	1.57
95%	1.233	1.574	0.088	0.086	0.040	0.043	1.612	1.63
99%	2.916	5.122	0.094	0.091	0.053	0.060	1.680	1.76
Minimum	0.001	0.000	0.018	0.014	0.003	0.000	1.102	1.11
Maximum	6.045	7.072	0.099	0.098	0.068	0.068	1.797	1.95

Observed, n = 1092; generated, n = 1000.

texture classifications. Joint multivariate distributions that incorporated correlations among hydraulic variables were developed for each class using an extensive soils database. The marginal distributions used in fitting these empirical data were selected as members of a family of distributions. Application of appropriate transformations resulted in variables that were approximately normally distributed, so that a multivariate normal distribution could be used to represent each of the joint density functions.

REFERENCES

Baes, C. F., III, and R. D. Sharp, A proposal for estimation of soil leaching and leaching constants for use in assessment models, *J. Environ. Qual.*, 12(1), 17–28, 1983.
Bresler, E., and G. Dagan, Convective and pore scale dispersive solute transport in unsaturated heterogeneous fields, *Water Resour. Res.*, 17(6), 1683–1693, 1981.
Brooks, R. H., and A. T. Corey, Hydraulic properties of porous media, *Hydrol. Pap. 3*, Colo. State Univ., Fort Collins, 1964.
Campbell, G. S., A simple method for determining unsaturated con-
ductivity from moisture retention data, *Soil Sci.*, 117(2), 311–314, 1974.
Carsel, R. F., R. S. Parrish, R. L. Jones, J. L. Hansen, and R. L. Lamb, Characterizing the uncertainty of pesticide movement in agricultural soils, *J. Contam. Hydrol*, in press, 1988.
Cox, P. C., and P. Baybutt, Methods for uncertainty analysis: A comparative survey, *Risk Anal.*, 1(2), 252–258, 1981.
El-Kadi, A. I., On estimating the hydraulic properties of soil, 2, A new empirical equation for estimating hydraulic conductivity for sands, *Adv. Water Resour.*, 8, 148–153, 1981.
El-Kadi, A. I., Variability of infiltration under uncertainty in unsaturated zone parameters, *J. Hydrol*, 90(1), 61–80, 1987.
Free, G. R., G. M. Browning, and G. W. Musgrave, Relative infiltration and related physical characteristics of certain soils, *Tech. Bull.* 279, U.S. Dep. of Agric., Washington, D.C., 1940.
Hillel, D., *Introduction to Soil Physics*, Academic, Orlando, Fla., 1982.
Holtan, H. N., C. B. England, G. P. Lawless, and G. A. Schumaker, Moisture-tension data for selected soils on experimental watersheds, *Rep. ARS 41-144*, U.S. Dep. of Agric., Washington, D.C., 1968.
Johnson, M. E., *Multivariate Statistical Simulation*, 230 pp., John Wiley, New York, 1987.
Johnson, N. L., and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, vol. 1, Houghton Mifflin Company, Boston, Mass., 1970.
Jury, W. A., Simulation on solute transport using a transfer function model, *Water Resour. Res.*, 18(2), 363–368, 1982.
Jury, W. A., Spatial variability of soil and physical parameters in

solute migration: A critical literature review, *Rep. EPRI EA-4228 RP2485-6*, Electr. Power Res. Inst., Palo Alto, Calif., 1985.

Kennedy, W. J., and J. E. Gentle, *Statistical Computing*, Marcel Dekker, New York, 1980.

Nielsen, D. R., J. W. Biggar, and K. T. Erh, Spatial variability of field measured soil-water properties, *Hilgardia, 42*(2), 215–259, 1973.

Rawls, W. J., and D. L. Brankensiek, Prediction of soil water properties for hydrologic modelling, in *Proceedings of Symposium on Watershed Management*, pp. 293–299, American Society of Civil Engineers, New York, 1985.

Rawls, W. J., D. L. Brankensiek, and K. E. Saxton, Estimation of soil water properties, *Trans. ASAE, 25*(5), 1316–1320, 1982.

Smith, C. N., R. S. Parrish, and R. F. Carsel, Estimating sample requirements for field evaluations of pesticide leaching, *J. Environ. Toxicol. Chem., 6*(5), 343–357, 1987.

Sharma, M. L., and A. S. Rogowski, Hydrological characterization of watersheds, in *Proceedings of the Natural Resources Modeling Symposium*, pp. 291–295, United States Department of Agriculture, Pingree Park, Colo., 1983.

Van Genuchten, M. Th., A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. J., 44*(5), 892–898, 1976.

Warrick, A. W., and D. R. Nielson, Spatial variability of soil physical properties in the field, in *Applications of Soil Physics*, edited by D. Hillel, pp. 319–344, Academic, Orlando, Fla., 1980.

R. F. Carsel, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA 30613.

R. S. Parrish, Computer Sciences Corporation, U.S. Environmental Protection Agency, Athens, GA 30613.