Development of Property-Transfer Models for Estimating the Hydraulic Properties of Deep Sediments at the Idaho National Engineering and Environmental Laboratory, Idaho

Scientific Investigations Report 2005–5114

U.S. Department of the Interior
U.S. Geological Survey

Development of Property-Transfer Models for Estimating the Hydraulic Properties of Deep Sediments at the Idaho National Engineering and Environmental Laboratory, Idaho

By Kari A. Winfield

Scientific Investigations Report 2005-5114

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2005

Contents

Figures

Tables

Conversion Factors, Datums, and Acronyms

Conversion Factors

Multiply	By	To obtain
centimeter (cm)	0.3937	inch
centimeter of water (cm-water)	0.01419	pound per square inch
centimeter per second (cm/s)	0.03281	foot per second
cubic centimeter (cm^3)	0.06102	cubic inch
gram (g)	0.03527	ounce, avoirdupois
gram per cubic centimeter (g/cm^3)	62.4220	pound per cubic foot
kilometer (km)	0.6214	mile
meter (m)	3.281	foot
millimeter (mm)	0.03937	inch
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F=(1.8×°C)+32$$

Datums

Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD of 1929).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD83).

Altitude, as used in this report, refers to distance above the vertical datum.

Acronyms

ICPP   Idaho Chemical Processing Plant

INEEL   Idaho National Engineering and Environmental Laboratory

INTEC   Idaho Nuclear Technology and Engineering Center

PPCC   probability-plot correlation coefficient

PTM   property-transfer model

RMSE   root-mean-square error

RWMC   Radioactive Waste Management Complex

SDA   Subsurface Disposal Area

SRP   Snake River Plain

TAN   Test Area North

TRA   Test Reactor Area

TSA   Transuranic Storage Area

VIF   variance inflation factor

VZRP   Vadose Zone Research Park

USDA   U.S. Department of Agriculture

USGS   U.S. Geological Survey

Development of Property-Transfer Models for Estimating the Hydraulic Properties of Deep Sediments at the Idaho National Engineering and Environmental Laboratory, Idaho

by Kari A. Winfield

Abstract

Because characterizing the unsaturated hydraulic properties of sediments over large areas or depths is costly and time consuming, development of models that predict these properties from more easily measured bulk-physical properties is desirable. At the Idaho National Engineering and Environmental Laboratory, the unsaturated zone is composed of thick basalt flow sequences interbedded with thinner sedimentary layers. Determining the unsaturated hydraulic properties of sedimentary layers is one step in understanding water flow and solute transport processes through this complex unsaturated system. Multiple linear regression was used to construct simple property-transfer models for estimating the water-retention curve and saturated hydraulic conductivity of deep sediments at the Idaho National Engineering and Environmental Laboratory. The regression models were developed from 109 core sample subsets with laboratory measurements of hydraulic and bulk-physical properties. The core samples were collected at depths of 9 to 175 meters at two facilities within the southwestern portion of the Idaho National Engineering and Environmental Laboratory—the Radioactive Waste Management Complex, and the Vadose Zone Research Park southwest of the Idaho Nuclear Technology and Engineering Center. Four regression models were developed using bulk-physical property measurements (bulk density, particle density, and particle size) as the potential explanatory variables. Three representations of the particle-size distribution were compared: (1) textural-class percentages (gravel, sand, silt, and clay), (2) geometric statistics (mean and standard deviation), and (3) graphical statistics (median and uniformity coefficient). The four response variables, estimated from linear combinations of the bulk-physical properties, included saturated hydraulic conductivity and three parameters that define the water-retention curve.

For each core sample, values of each water-retention parameter were estimated from the appropriate regression equation and used to calculate an estimated water-retention curve. The degree to which the estimated curve approximated the measured curve was quantified using a goodness-of-fit indicator, the root-mean-square error. Comparison of the root-mean-square-error distributions for each alternative particle-size model showed that the estimated water-retention curves were insensitive to the way the particle-size distribution was represented. Bulk density, the median particle diameter, and the uniformity coefficient were chosen as input parameters for the final models. The property-transfer models developed in this study allow easy determination of hydraulic properties without need for their direct measurement. Additionally, the models provide the basis for development of theoretical models that rely on physical relationships between the pore-size distribution and the bulk-physical properties of the media. With this adaptation, the property-transfer models should have greater application throughout the Idaho National Engineering and Environmental Laboratory and other geographic locations.

Introduction

Flow processes within the highly stratified unsaturated zone at Idaho National Engineering and Environmental Laboratory (INEEL) are not well understood. The thick basalt layers are fractured to massive, contain rubble zones, and are interbedded with thinner sedimentary layers. All these features potentially cause preferential contaminant and water flow through the unsaturated zone. To better understand these processes, the role that sedimentary interbeds and surficial sediments play in unsaturated flow needs to be determined; this typically is accomplished using flow simulations. Flow models require input of basic unsaturated hydraulic properties, such as water-retention and unsaturated hydraulic conductivity curves, which are representative of the modeled media. The water-retention curve ($\theta(\psi)$) is a measure of how the water content (θ) varies with the negative water (or matric) pressure (ψ) of a particular soil or sedimentary material. Unsaturated hydraulic conductivity (K), or the rate at which water is transmitted through a sedimentary material, varies with θ or ψ. Saturated hydraulic conductivity (K_{sat}) is the flow rate at the maximum degree of saturation (θ_{sat}). Knowledge of K_{sat} is needed to estimate $K(\theta)$ from established models. K_{sat} often is measured in place of $K(\theta)$ because K_{sat} is easier to determine

than $K(\theta)$, and because it can be used along with $\theta(\psi)$ to estimate $K(\theta)$ using parametric models (for example, Mualem, 1976). Field tests can provide information about the average unsaturated hydraulic properties over a given region at depth, but when studying sedimentary interbeds this task becomes more challenging because of the overlying and underlying fractured basalt units. Alternatively, minimally disturbed core samples can be collected by drilling deep boreholes, and their hydraulic properties can be determined in a laboratory; however, this process is costly and time consuming. Drilling can require months to complete and obtaining undisturbed core samples can be problematic, especially in coarse-textured materials. Laboratory measurement of unsaturated hydraulic properties requires specialized equipment and expertise, in addition to requiring several weeks to complete measurements for a single core sample. These difficulties are compounded at the INEEL by the thick vadose zone (up to 200 m), the complex stratigraphy, and the large site area.

An alternative to measuring the unsaturated hydraulic properties of core samples in the laboratory is to develop a property-transfer model (PTM) that can predict these properties from more easily measured bulk-physical properties, such as bulk density (ρ_{bulk}), particle density (ρ_{part}), and particle-size distribution. General purpose models of this type exist, but they are inadequate for sediments at the INEEL because they were not developed from measurements specific to the site. Hydraulic PTMs have not been developed for the sedimentary interbeds at the INEEL. These types of models will be a cost effective way to estimate unsaturated hydraulic properties for use in flow models, to predict contaminant travel times and flow paths, and to provide greater understanding of unsaturated flow processes in the complex unsaturated zone at the INEEL. Property-transfer modeling is a useful approach to estimating hydraulic properties, even when the estimates have a large margin of error, because hydraulic-property measurements cannot be made for every point in space.

Often when core samples are not obtained during drilling, borehole cuttings are collected that can be analyzed for particle size. The majority of laboratory data available at the INEEL consists of particle-size distributions and other bulk-physical-property measurements obtained from borehole cuttings. The U.S. Geological Survey (USGS), Bechtel BWXT Idaho, LLC, previous contractors, and other institutions have investigated hydrologic processes within the vadose zone of the INEEL for several decades; however, relatively little work has been done in the laboratory to characterize the hydraulic properties of sedimentary interbeds. Unsaturated hydraulic property estimates by PTMs can save considerable time and effort by reducing the need to measure unsaturated hydraulic properties in the laboratory. Additionally, estimation of hydraulic properties will enable a broader understanding of the distribution of hydraulic and bulk-physical properties at the INEEL, which is essential in developing models of ground-water flow and contaminant transport.

Site Background

The INEEL was established in 1949 under the U.S. Department of Energy (formerly, the U.S. Atomic Energy Commission) for nuclear energy research. The site hosts several facilities (fig. 1) at least four of which have generated, stored, or disposed of radioactive, organic, and inorganic wastes. These include the (1) Radioactive Waste Management Complex (RWMC), (2) Idaho Nuclear Technology and Engineering Center (INTEC), formerly known as the Idaho Chemical Processing Plant (ICPP), (3) Test Reactor Area (TRA), and (4) Test Area North (TAN). Because of the site's history of chemical-waste production and ground-water resource contamination, vadose-zone investigations are used to determine or forecast critical pathways for contaminant migration to the Snake River Plain (SRP) aquifer, and to understand flow and transport processes within the geologically complex subsurface. Because hydraulic property measurements are available only on core samples collected from the Vadose Zone Research Park (VZRP) and RWMC, this report focuses on these two locations within the INEEL.

The RWMC (figs. 1 and 2), which covers approximately 0.6 km^2, was established in 1952 as an area for controlled management of solid radioactive waste produced on site. Facilities include the Subsurface Disposal Area (SDA) for shallow burial of mixed low-level wastes and the Transuranic Storage Area (TSA) for above-ground waste storage containing elements with atomic numbers equal to or higher than uranium. Mixed transuranic and low-level wastes were buried in pits and trenches at the SDA from 1952 to 1970. Since 1970, transuranic wastes have been stored in above-ground containers on asphalt pads at the TSA, and low-level waste burial has continued at the SDA (Anderson and Lewis, 1989).

The INTEC is in the southwestern part of the INEEL covering an area of about 0.8 km^2 (figs. 1 and 2). The INTEC was established in the 1950's for reprocessing spent nuclear fuel from government reactors. The facility houses an underground tank farm for storing high-level wastes, and above-ground storage areas for spent nuclear wastes (Bechtel BWXT, LLC, 2003). Between 1952 and 1984, radiochemical and chemical wastes produced on site were disposed of in a 183-m-deep injection well, which penetrated the SRP aquifer. Wastes produced after 1984 were disposed of in two percolation ponds (fig. 3) directly south of the facility (Cecil and others, 1991; Bartholomay and others, 1997; Bartholomay, 1998). Contaminants detected in perched water bodies beneath these ponds led to their discontinued use in 1995. In 2001, new INTEC percolation ponds were constructed within the 3-km^2 VZRP (fig. 3), about 9.5 km southwest of the original ponds.
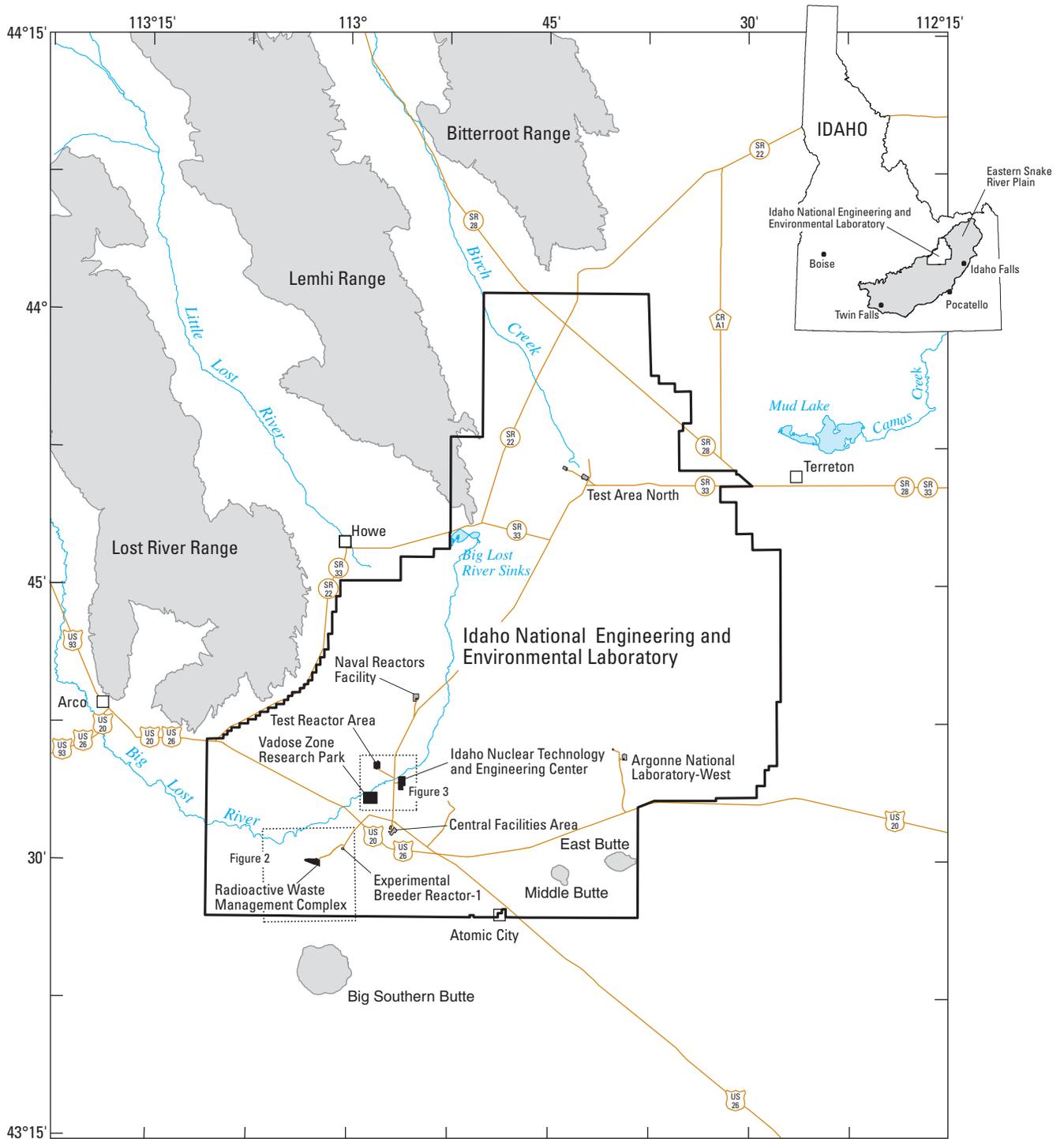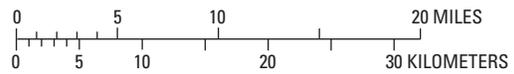
Figure 1. Location of the Idaho National Engineering Environmental Laboratory and selected facilities, Idaho.
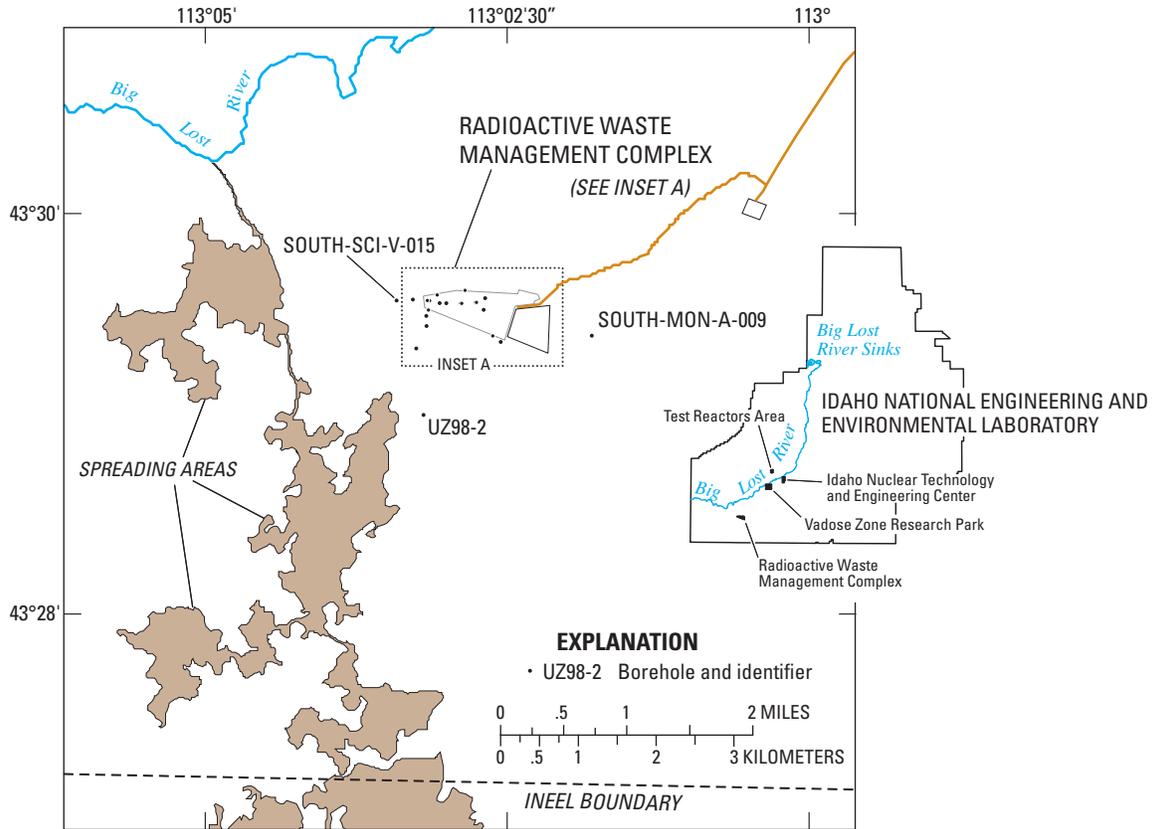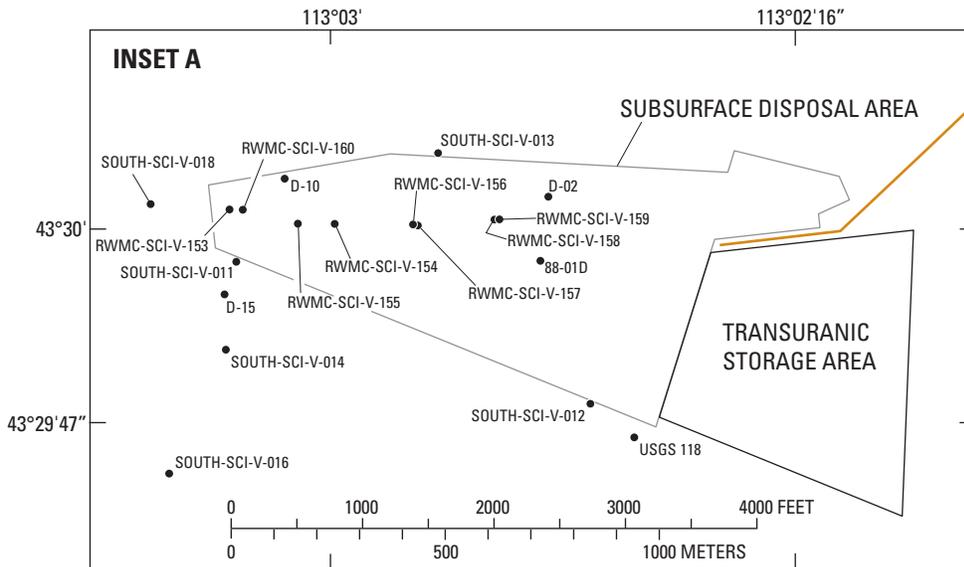
Figure 2.   Boreholes in the Radioactive Waste Management Complex, Idaho National Engineering and Environmental Laboratory, Idaho, where measurements on sedimentary interbed core samples were used to develop hydraulic property-transfer models.
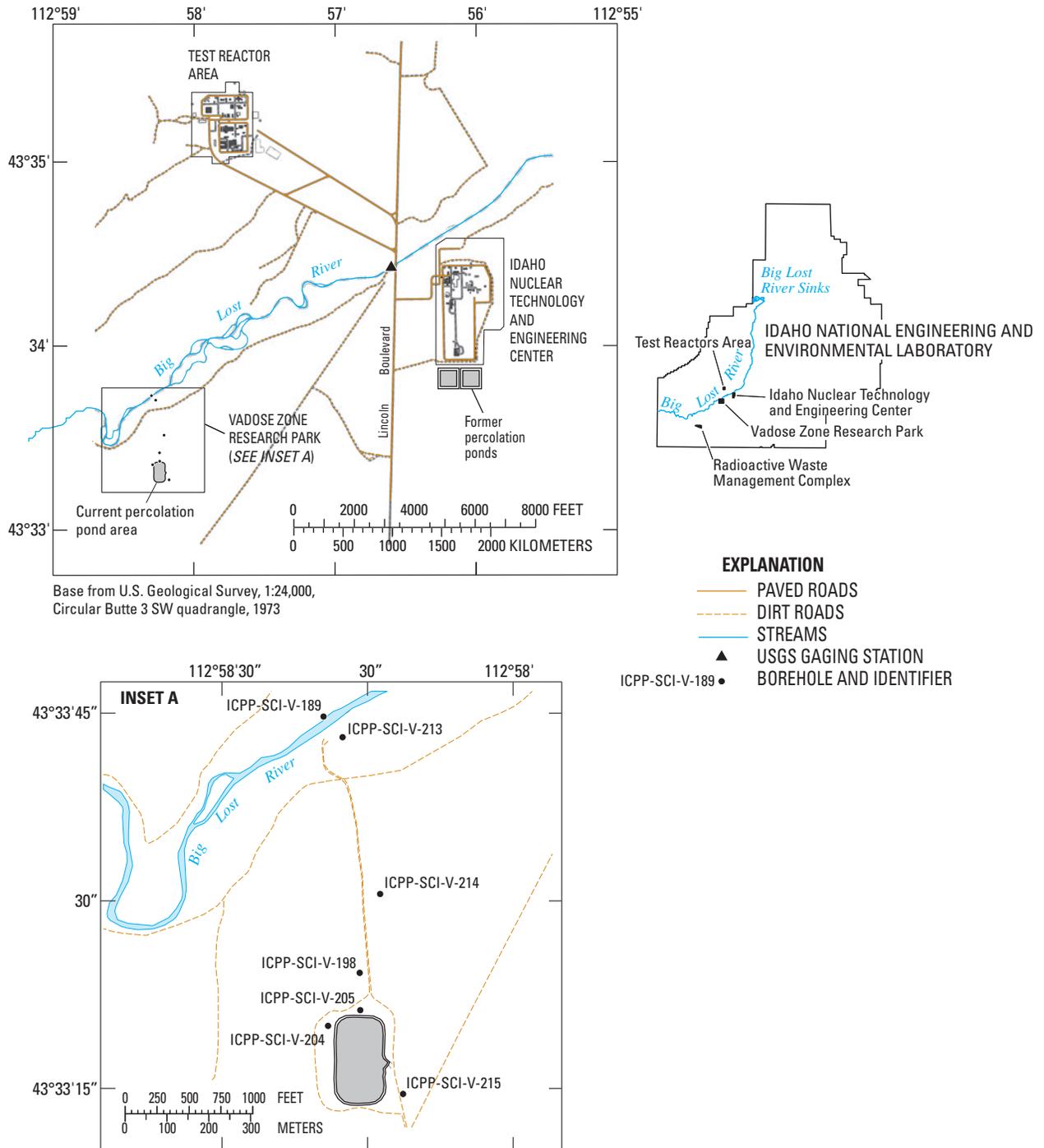
Figure 3. Boreholes in the Vadose Zone Research Park, Idaho National Engineering and Environmental Laboratory, Idaho, where measurements on sedimentary interbed core samples were used to develop hydraulic property-transfer models.

Geohydrologic Setting

The INEEL occupies about 2,300 km^2 within the eastern SRP, a northern extension of the Basin and Range Province. The eastern SRP is a northeast trending basin, approximately 320 km long and 80 to 110 km wide, which slopes gently to the southwest and is bordered by northwest trending mountain ranges.

The eastern SRP is underlain by interbedded volcanic and sedimentary units extending up to 3,000 m below the surface. The sedimentary layers represent quiet intervals between volcanic eruptions and are of eolian, fluvial, and lacustrine origin with large amounts of sand, silt, and clay. Volcanic units, composed primarily of basalt flows, welded ash flows, and rhyolite, may be vesicular to massive with either horizontal or vertical fracture patterns. Near the INTEC facility, boreholes drilled to 200-m depths penetrate a sequence of 23 basalt-flow groups and 15 to 20 sedimentary interbeds (Anderson, 1991). The surficial sediments near the INTEC consist of gravelly alluvium, range from 2- to 20-m thick, and are thickest to the northwest (Anderson and others, 1996). Beneath the RWMC, 10 basalt-flow groups and 7 major sedimentary interbeds have been identified from boreholes drilled up to 220-m depth. Surficial sediments are composed primarily of silt and clay, and range in thickness from 0 to 7 m (Rightmire and Lewis, 1987; Anderson and Lewis, 1989).

The climate of the eastern SRP is semi-arid with an average annual precipitation of 22 cm. Parts of the SRP aquifer underlie the INEEL. The depth to the water table ranges from 60 m in the northern part of the INEEL to about 200 m toward the south (Barraclough and others, 1981; Liszewski and Mann, 1992). The depth to the water table is about 145 m below the INTEC and about 180 m below the RWMC. The predominant direction of ground-water flow is from northeast to southwest. Recharge to the aquifer primarily is from irrigation water diversions from streams, precipitation and snowmelt, underflow from tributary-valley streams, and seepage from surface water bodies (Hackett and others, 1986). Within the INEEL boundaries, the Big Lost River (fig. 1) is an ephemeral stream, which flows from southwest to northeast about 3 km north of the RWMC (fig. 2) and less than 1 km north of the INTEC (fig. 3). Because of the proximity of the Big Lost River to waste disposal and storage facilities, a diversion dam was constructed in 1958 upstream of the SDA to reduce the threat of flooding (Barraclough and others, 1967). During high flow periods, flow is diverted to topographic depressions (referred to as spreading areas) less than 2 km west of the RWMC (fig. 2).

Previous Investigations

This section describes studies in which hydraulic property measurements specific to the INEEL site were measured and gives an overview of PTMs. Few studies have been done in which the hydraulic properties of sedimentary interbeds have been determined on core samples. Only two locations within the INEEL have been characterized, the RWMC and the VZRP. Two classes of PTMs exist in the literature, those based on empirical relationships and those based on theoretical ones. Within the empirical model category, multiple linear regression often is used as a tool for model development.

INEEL Hydraulic Property Measurements

Most studies of contaminant transport and measurement of hydraulic properties at the INEEL have been done near the RWMC (fig. 1), where potential migration of shallow, buried hazardous waste through the unsaturated zone poses concern for water quality in the SRP aquifer. McElroy and Hubbell (1990) presented measurements of $\theta(\psi)$, particle size, and ρ_{bulk} on intact sedimentary core samples collected from eight boreholes near the RWMC. Core samples from sedimentary interbeds were collected from five of these boreholes (USGS 118, 88-01D, D-02, D-10, and D-15; shown in fig. 2). Deep core-sample measurements of $\theta(\psi)$ and $K(\theta)$, in addition to bulk-physical properties, were presented by Perkins and Nimmo (2000) for borehole UZ98-2, about 1.5 km southwest of the SDA (fig. 2). Measurements from an additional 16 boreholes near the SDA also have been completed on sediment core samples collected from depths of 29.45 to 78.02 m below land surface (S. Magnuson, written commun., 2002; fig. 2 and table 1). Perkins (2003) and Winfield (2003) presented unsaturated hydraulic- and bulk-physical-property measurements for seven boreholes near the current percolation pond area for the INTEC facility at the VZRP (fig. 3).

Several studies have been done to characterize the unsaturated hydraulic properties of surficial sediments, defined as the sediment and soil overlying the shallowest basalt layer (Barraclough and others, 1976; McElroy and Hubbell, 1990; Borghese, 1991; Shakofsky, 1995; Nimmo and others, 1999). Measurements on surficial soils and sediments are not considered in the models presented in this report because their hydraulic properties are affected by structural features arising from plant growth, compaction by anthropogenic sources, animal burrows, and shrink-swell phenomena. Nimmo and others (2004), table 1, present a more complete summary of available INEEL measurements.

Table 1.   Sedimentary interbed core samples with hydraulic and bulk-physical property measurements at the Idaho National Engineering and Environmental Laboratory (INEEL), Idaho.

[Water retention is defined as water content (θ) as a function of matric pressure (ψ). **Abbreviations:** m, meter; RWMC, Radioactive Waste Management Complex; VZRP, Vadose Zone Research Park]

Borehole identification No.	Borehole location	Depth interval (m)	Number and type of core sample measurements					
			Saturated hydraulic conductivity (K_{sat})	Unsaturated hydraulic conductivity as a function of water content ($K(\theta)$)	Water retention ($\theta(\psi)$)	Bulk density (ρ_{bulk})	Particle density (ρ_{part})	Particle-size distribution
Magnuson, Swen, Bechtel BWXT Idaho, LLC (written commun., 2002)								
RWMC-SCI-V-153, RWMC-SCI-V-154, RWMC-SCI-V-155, RWMC-SCI-V-156, RWMC-SCI-V-157, RWMC-SCI-V-158, RWMC-SCI-V-159, RWMC-SCI-V-160, SOUTH-MON-A-009, SOUTH-SCI-V-011, SOUTH-SCI-V-012, SOUTH-SCI-V-013, SOUTH-SCI-V-014, SOUTH-SCI-V-015, SOUTH-SCI-V-016, SOUTH-SCI-V-018	RWMC	29.45–78.02	41	0	51	51	51	51
McElroy and Hubbell (1990)								
88-01D, D-02, D-10, D-15, USGS 118	RWMC	9.60–17.48	16	0	16	16	16	[1]16
Perkins and Nimmo (2000)								
UZ98-2	RWMC	42.98–50.30	18	18	18	18	18	18
Perkins (2003)								
ICPP-SCI-V-215	VZRP	45.54–59.92	12	12	14	14	14	14
Winfield (2003)								
ICPP-SCI-V-189, ICPP-SCI-V-198, ICPP-SCI-V-204, ICPP-SCI-V-205, ICPP-SCI-V-213, ICPP-SCI-V-214	VZRP	36.59–56.33	10	10	10	10	10	10
Number of samples, *n*			97	40	109	109	109	[1]109

[1]Partial particle-size data were reported for sample USGS 118 at 173.48 to 174.39 m from McElroy and Hubbell (1990).

Overview of Property-Transfer Models

PTMs can be classified into two groups: (1) empirical models, which rely on statistical methods to determine patterns in the data, and (2) quasi-physical models, which are based on theoretical or physical relationships between pore sizes and particle or aggregate sizes (Arya and Paris, 1981; Haverkamp and Parlange, 1986; Nimmo, 1997; Haverkamp and Reggiani, 2002). Methods for predicting $\theta(\psi)$ are described here, however, these methods also can be applied to prediction of $K(\theta)$.

Empirical models typically involve multiple linear-regression analyses (although some use a neural network procedure) and can be further subdivided based on the specific approach used to estimate unsaturated hydraulic-property curves. One approach involves fitting a parametric $\theta(\psi)$ function (or set of functions) to the $\theta(\psi)$ measurements, and developing separate regression equations for each $\theta(\psi)$ parameter (Campbell, 1985; Saxton and others, 1986; Wösten and van Genuchten, 1988; Vereecken and others, 1989; Campbell and Shiozawa, 1992; Schaap and others, 1998, 2001). Another approach involves developing unique equations for θ at the values of ψ determined during measurement of $\theta(\psi)$ (Gupta and Larson, 1979; Rawls and Brakensiek, 1982; Puckett and others, 1985; Mecke and others, 2002). A third approach uses at least one measured value of $\theta(\psi)$, in addition to bulk-physical properties, as input (Gregson and others, 1987; Schaap and others, 1998).

Several studies also have predicted K_{sat} using multiple linear regression (Puckett and others, 1985; Campbell and Shiozawa, 1992; Jabro, 1992; Mecke and others, 2002) or neural networks (Schaap and others, 1998, 2001). By empirical modeling, several equations are required to completely predict $\theta(\psi)$ because θ varies uniquely in relation to ψ. Because K_{sat} relates only to one θ value, θ_{sat}, a single equation can be used to directly relate K_{sat} to a set of bulk-physical properties using statistical techniques. The remainder of the $K(\theta)$ curve can be predicted using any of the aforementioned techniques, providing measurements of $K(\theta)$ are available. Alternatively, estimates of K_{sat} and $\theta(\psi)$ obtained by PTMs can be used to predict $K(\theta)$ using Mualem's (1976) capillary bundle model.

Most PTMs available in the literature were developed for specific applications, such as for widespread use (Schaap and others, 1998, 2001) or for a specific site or soil type (Jabro, 1992; Puckett and others, 1985; Mecke and others, 2002). PTMs developed for use in agricultural studies typically are calibrated using repacked samples where gravels have been removed. Mecke and others (2002) reaffirmed the conclusion of Puckett and others (1985) and Williams and others (1992) that a predictive model should be developed on materials of similar mineralogy and genesis. PTMs developed for a specific site cannot describe the diverse range of soil textures and structures found at other sites. Factors that may cause PTMs to differ between sites include water chemistry, mineralogy of sediment, liquid-solid interactions, colloidal transport and potential blockage of pores, development of cements in pores, and sediment structure (macropores, grain shape, orientation, roundness, and packing arrangement).

Most PTMs available in the literature have been developed for surficial soils, often with characteristics particular to a given site and soil type, such as Lower Coastal Plain ultisols (Puckett and others, 1985) or glacially-derived podzols (Mecke and others, 2002). The development of reliable PTMs for deeper sediments (below the zone of soil development) is needed, especially as the USGS and other agencies move toward characterizing larger portions of the subsurface for studies of ground-water recharge. Other PTMs calibrated using large databases of soil properties are inappropriate for the INEEL because the databases often contain too diverse a mixture of samples with conflicting characteristics, such as repacked samples included with intact samples, or samples with pedologic structure included with single-grained soils.

Purpose and Scope

This report describes the development of PTMs for predicting $\theta(\psi)$ and K_{sat} of the sedimentary interbeds at the INEEL. The PTMs are constrained by available data to two locations within the INEEL, the RWMC and the VZRP.

Because surficial sediments possess different physical characteristics than the sedimentary interbeds, these data were not used to develop the PTMs in this study. Multiple linear-regression techniques were used to estimate hydraulic properties from bulk-physical properties based on 109 sets of core-sample measurements compiled from previous studies (S. Magnuson, written commun., 2002; McElroy and Hubbell, 1990; Perkins and Nimmo, 2000; Perkins, 2003; Winfield, 2003; see table 1). Three regression equations were developed for parameters defining the $\theta(\psi)$ curve (θ_{sat} and two parameters from the Rossi-Nimmo (1994) junction model). The $\theta(\psi)$ parameters and K_{sat} were used as the dependent variables in the regression models. Parameters describing the particle-size distribution, ρ_{bulk}, and ρ_{part} values were used as potential independent variables. This report describes the available data, data processing, multiple linear-regression assumptions and approach, and regression equations developed for each hydraulic parameter. Three representations of particle-size data were compared to determine the optimal regression variate to use in the final PTMs. Selection of the best PTM was based on ease of calculating particle-size parameters, model behavior (in terms of meeting the assumptions of multiple linear regression), model fit, and potential for developing a theoretical model. The results of this report contribute to the characterization of the unsaturated zone at the INEEL by providing a tool that can be used to predict hydraulic properties required as input to unsaturated-flow simulations.

Property-Transfer Model Calibration and Approach

To develop PTMs for the INEEL sedimentary interbeds, multiple linear-regression equations were used to estimate hydraulic properties (measurements of θ_{sat} and K_{sat}, and curve-fit parameters for $\theta(\psi)$) from linear combinations of bulk-physical properties (ρ_{bulk}, ρ_{part}, and particle-size parameters). To estimate the entire $\theta(\psi)$ curve from saturation to oven dryness, the Rossi-Nimmo (1994) junction model was used to characterize the $\theta(\psi)$ measurements. Multiple linear-regression equations then were developed for each of the three parameters defining the Rossi-Nimmo (1994) junction model. Three alternative models were compared that used different representations of the particle-size distribution as input, along with ρ_{bulk} and ρ_{part}: (1) model A included textural-class percentages, (2) model B included geometric statistical parameters, and (3) model C included graphical statistical parameters. The basic steps of the multiple linear-regression procedure used in this study are outlined in figure 4 for $\theta(\psi)$. The same steps were followed for K_{sat}, except no curve-fitting was involved.

RAW DATA   **DATA PROCESSING**   **REGRESSION MODEL INPUT**   **REGRESSION MODEL OUTPUT**

HYDRAULIC PROPERTIES
(RESPONSE VARIABLES, Y)

Water-retention measurements → Rossi-Nimmo (1994) junction model, curve fit

Saturated water content (θ_{sat})

OR

Scaling parameter for matric pressure (ψ_o)

OR

Curve-shape parameter (λ)

EACH Y VARIABLE

BULK PHYSICAL PROPERTIES
(EXPLANATORY VARIABLES, X)

Particle-size measurements → Frequency of particles per bin

Cumulative frequency distribution

MODEL B:
Geometric statistics (mean and standard deviation)

OR

MODEL A:
USDA textural-class percentages

OR

MODEL C:
Graphical statistics (median and uniformity coefficient)

EACH SET OF X VARIABLES

AND

Bulk-density and particle-density measurements

no processing

Bulk-density and particle- density measurements

Least-squares multiple linear regression

9 separate equations

Coefficients for each y and set of x variables, goodness-of-fit values

Figure 4.   Steps used to develop multiple linear-regression models for water-retention parameters.

Data Sets and Measurement Techniques

The hydraulic- and bulk-physical-property data sets used to develop the $\theta(\psi)$ and K_{sat} models for the INEEL sedimentary interbeds, including the number and type of core-sample measurements, are summarized in table 1. These data represent the potential number of core samples for developing the $\theta(\psi)$ and K_{sat} PTMs; smaller subsets were used to develop the final regression models after evaluating the data for errors and missing values. To develop the multiple linear-regression equations, complete data sets of hydraulic ($\theta(\psi)$ or K_{sat}) and bulk-physical (ρ_{bulk}, ρ_{part}, and particle-size distribution) properties were needed for each core sample. Core samples with missing values of either the hydraulic or bulk-physical properties were removed from the regression analyses. The particle-size distributions needed to have sufficient resolution, including points in the clay-sized (less than 0.002 mm) fraction, to compute geometric particle-size statistics, textural-class percentages, or particle sizes at particular cumulative percentiles (10th, 50th, and 60th). Because hydraulic property measurements are only available for the RWMC and VZRP, the PTMs presented in this report can only be generalized to interbeds in the southwestern portion of the INEEL. Surficial sediments were excluded from the PTM development because of likely physical differences from the interbed sediments caused by soil development processes, freeze-thaw phenomena, or animal or microbial activity. Five data sets were used in model development:

1. RWMC, 51 core samples of Bechtel BWXT Idaho, LLC (S. Magnuson, written commun., 2002),

2. RWMC, 16 core samples of McElroy and Hubbell (1990),

3. RWMC, 18 core samples of Perkins and Nimmo (2000),

4. VZRP, 14 samples of Perkins (2003), and

5. VZRP, 10 samples of Winfield (2003).

Boreholes where core samples were collected are shown in figure 2 for the RWMC and in figure 3 for the VZRP. The McElroy and Hubbell (1990) data were measured by Daniel B. Stephens and Associates, Inc., New Mexico. The data of Magnuson (written commun., 2002) were measured by the Southwest Research Institute, Texas. The data of Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003) were measured by the USGS. Partial particle-size data were reported for one McElroy and Hubbell (1990) sample from borehole USGS 118 at 173.48 to 174.39 m. Ten samples from Magnuson (written commun., 2002) and two samples from Perkins (2003) did not have K_{sat} measurements.

Techniques used to determine unsaturated hydraulic properties of the 109 core samples differed among studies. All $\theta(\psi)$ measurements were determined on minimally disturbed samples except for those of Magnuson (written commun., 2002). These samples were sieved to remove the gravel (> 2 mm) fraction and then repacked. For the Magnuson (written commun., 2002) data, five θ points were determined per sample, at ψ values of -200, -400, -600, -800, and -1,000 cm-water, using the pressure-plate technique. For the McElroy and Hubbell (1990) data, eight points on the $\theta(\psi)$ curve were determined per sample, spanning the range of ψ from about -50 to -15,000 cm-water. Points in the wet range (> -400 cm-water) were determined by the hanging column method, and drier points were determined by the pressure-plate technique. The $\theta(\psi)$ measurements of Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003) were completed as part of the steady-state centrifuge method (Nimmo and others, 1994; Conca and Wright, 1998; Nimmo and others, 2002), which was used to determine $K(\theta)$. Tensiometers were used to determine equilibrium ψ at different θ values. When samples were too dry to use with a tensiometer, the filter paper method (Fawcett and Collis-George, 1967; Greacen and others, 1987) was used. For the Perkins and Nimmo (2000) samples, additional dry-range $\theta(\psi)$ points were determined for the range of ψ between -10^5 and -10^7 cm-water using a chilled mirror hygrometer (Winfield, written commun., 2001). The number of $\theta(\psi)$ points for these studies varied from 7 to 18, with an average of 14 points for Perkins and Nimmo (2000), 10 points for Perkins (2003), and 9 points for Winfield (2003).

Degree of saturation was either measured directly or inferred from other properties. For the Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003) studies, saturation was achieved by submerging the core samples in a wetting solution up to the sample height. To determine θ_{sat} values, the samples were weighed after measurement of K_{sat} and the weights were converted to volumetric θ values using ρ_{bulk}. For the McElroy and Hubbell samples, θ_{sat} was determined prior to beginning $\theta(\psi)$ and K_{sat} measurements, after vacuum saturating the samples. θ_{sat} was not measured for the Magnuson samples. In this report, the maximum water content for the Magnuson samples was assumed to equal the total porosity determined from the repacked sample ρ_{bulk}.

For all studies, K_{sat} was measured on intact core samples. K_{sat} values for the McElroy and Hubbell (1990) samples were determined using either the falling-head or the constant-head method. Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003) used either the traditional falling-head method, or the falling-head method adapted for use with a centrifuge (Nimmo and Mello, 1991; Nimmo and others, 2002) for low K_{sat} materials. K_{sat} values from Magnuson (written commun., 2002) were measured using the constant-head method.

Particle-size distributions were determined by sieve analysis, hydrometer, or laser light scattering. McElroy and Hubbell (1990) used sieve analysis for particle sizes greater than 0.075 mm, and the hydrometer method for sizes less than 0.075 mm. The number of points analyzed per sample was 20, with the exception of four samples that only had seven points greater than 0.075 mm. For one of these samples (USGS 118 at 173.48 to 174.39 m) the diameter at the 10th percentile (d_{10}), the geometric statistics, and textural-class percentages could not be calculated. For the Magnuson (written commun., 2002) samples, sieve analysis was used for particle sizes greater than 0.075 mm, and the hydrometer method was used for sizes less than 0.075 mm, with 20 points determined per sample. Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003) used sieve analysis for particles greater than 2 mm. A laser diffraction method was used for particle sizes from 4×10^{-5} mm to 2 mm, with 116 points analyzed.

Total porosity was calculated from measurements of ρ_{part} and ρ_{bulk}. Values of ρ_{part} were determined on small subsamples from the original core samples. The pycnometer method of Blake and Hartge (1986) was used to measure ρ_{part} on the core samples from Perkins and Nimmo (2000), Perkins (2003), and Winfield (2003). The entire particle-size distribution was analyzed for these samples. For the core samples of Magnuson (written commun., 2002) and McElroy and Hubbell (1990), ρ_{part} was determined using the standard pycnometer method D854-83 (now an archived method) of the American Society for Testing and Materials (1989). For the McElroy and Hubbell (1990) samples, material less than 4.75 mm was analyzed, while material less than 2 mm was analyzed for the Magnuson (written commun., 2002) samples. For the data set of Perkins and Nimmo (2000), the average ρ_{part} value of 2.65 g/cm^3 (measured for 4 core samples) was used for all 18 core samples. Values of ρ_{bulk} were determined on intact core samples for all data sets used in model calibration. Total porosity was calculated by $1 - \rho_{bulk}/\rho_{part}$. Although total porosity values were not used in developing the regression models, total porosity was used as a baseline for comparing what the core sample degree of saturation should be. Because total porosity is a measure of all void space in a sample, including any dead-end pores and the internal porosity of mineral grains, it should be greater than the measured θ_{sat} value (which can be considered a measure of effective porosity).

Water-Retention Curve-Fit Parameters

Because θ varies nonlinearly as a function of ψ, one or more functions involving multiple parameters are needed to describe the entire range of θ. One approach to developing a PTM for $\theta(\psi)$ involves developing multiple linear-regression equations for several θ values associated with particular ψ values. Difficulties applying this method to INEEL measurements included that (1) the number of measured $\theta(\psi)$ points varied between data sets, and (2) θ values were not

obtained at the same ψ values for all samples due to different methods used to measure $\theta(\psi)$. An alternate way to develop a PTM for $\theta(\psi)$ is to fit a curve to the measured $\theta(\psi)$ points for each core sample. This approach yields parameters that uniquely define each curve. Multiple linear regression is then used to develop equations relating each of these parameters to a set of bulk-physical properties.

The Rossi-Nimmo (1994) junction model (fig. 5) was chosen to fit the $\theta(\psi)$ measurements because this model is more physically realistic over the entire range of θ from saturation to oven dryness than other parametric models (Brooks and Corey, 1964; van Genuchten, 1980) that use residual water content (θ_r) as an optimized parameter; θ_r has no physical meaning. According to capillary theory the largest pores are associated with ψ values near zero and drain first, followed by drainage of successively smaller pores as θ approaches θ_r. The asymptotic approach to θ_r means that the number of small pores approaches infinity, which is physically unrealistic. The $\theta(\psi)$ curve represented by the Rossi-Nimmo (1994) junction model goes to zero θ at a fixed value of ψ calculated for the conditions of oven dryness (ψ_d). Another advantage to using the junction model is that it can be combined with the capillary-bundle model of Mualem (1976) to estimate $K(\theta)$ (Fayer and others, 1992; Rossi and Nimmo, 1994; Andraski, 1996; Andraski and Jacobson, 2000).
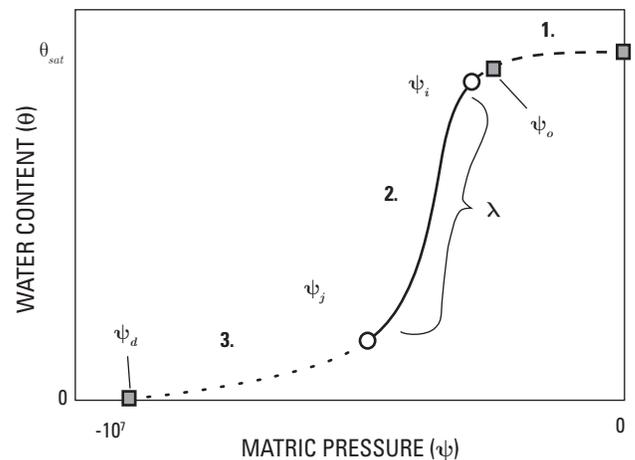
EXPLANATION

1.	Parabolic function, $\psi_i \leq \psi \leq 0$
2.	Power law function, $\psi_j \leq \psi \leq \psi_i$
3.	Logarithmic function, $\psi_d \leq \psi \leq \psi_j$
θ_{sat}	Saturated water content
ψ_o	Scaling parameter for matric pressure
λ	Curve-shape parameter
ψ_i	Junction point for **1.** and **2.**
ψ_j	Junction point for **2.** and **3.**
ψ_d	Oven-dry matric pressure

Figure 5. Example of water-retention ($\theta(\psi)$) curve showing components of the curve-fit model developed by Rossi and Nimmo (1994).

The Rossi-Nimmo (1994) junction model requires three parameters to describe the entire $\theta(\psi)$ curve, therefore, three separate multiple linear-regression equations are needed. By contrast, models that estimate θ at particular ψ values usually require 7 to 12 separate regression equations. These equations typically range from $\psi = 0$ at saturation to approximately -15,000 cm-water. For the abundance of silt-loam-textured samples found at the INEEL, -15,000 cm-water corresponds to saturations that vary from 23 to 88 percent, a range that does not adequately describe the dry end of the $\theta(\psi)$ curve. Using a $\theta(\psi)$ model that extends from 100 to 0 percent saturation avoids this problem. Additionally, developing multiple linear-regression equations for the $\theta(\psi)$ model parameters saves the extra step of having to fit the estimated $\theta(\psi)$ points in order to use these in unsaturated-flow simulations.

The Rossi-Nimmo (1994) junction model consists of three functions joined at two points (fig. 5):

1.   Parabolic function for the wet range of ψ,

2.   Power law function (Brooks and Corey, 1964) for the middle range of ψ, and

3.   Logarithmic function for the dry range of ψ.

This model has two independent parameters: (1) the scaling factor for ψ (ψ_o), and (2) the curve-shape parameter (λ). Sometimes, ψ_o is associated with ψ at which air first enters a porous material during desaturation (referred to as "air-entry pressure"), but, actually, air begins displacing water in the largest pores at a higher (less negative) ψ than ψ_o as evidenced by the departure of θ from saturation earlier than ψ_o on the $\theta(\psi)$ curve (fig. 5). In this study, the units of ψ_o are expressed in centimeters of water (cm-water). The curve-shape parameter λ indicates the relative steepness of the middle portion of the $\theta(\psi)$ curve, described by the power-law function. Larger λ values cause the drainage portion of the $\theta(\psi)$ curve to appear steeper. Unlike the model of Brooks and Corey, which holds θ fixed between $\psi = 0$ and the "air-entry pressure", the junction model produces a smooth curve near saturation, represented by a parabolic function, that allows the pore-size distribution (the first derivative of the $\theta(\psi)$ curve) to be represented more realistically.

The parabolic function applies for $\psi_i \leq \psi \leq 0$, and is represented by:

$$\frac{\theta}{\theta_{sat}} = 1 - c\left(\frac{\psi}{\psi_o}\right)^2, \tag{1}$$

where

θ_{sat} is expressed volumetrically and

c is a dimensionless constant calculated from an analytical function involving the parameter λ, which also is dimensionless.

The power law function applies for $\psi_j \leq \psi \leq \psi_i$, and is represented by:

$$\frac{\theta}{\theta_{sat}} = \left(\frac{\psi_o}{\psi}\right)^\lambda, \tag{2}$$

The logarithmic function applies for $\psi_d \leq \psi \leq \psi_j$, and is represented by:

$$\frac{\theta}{\theta_{sat}} = A\ln\left(\frac{\psi_d}{\psi}\right), \tag{3}$$

where A is the slope of the dry-range part of the $\theta(\psi)$ curve on a semilog scale, calculated from a combination of λ, ψ_o, and ψ_d. The junction points (ψ_i and ψ_j) are calculated from analytical functions of ψ_o and λ.

Optimized values of ψ_o and λ for each core sample were determined using a nonlinear regression program based on the modified Gauss-Newton least-squares method found in the Statistics Toolbox version 3 of MATLAB 6, Release 12 (The Mathworks, Inc., Natick, MA). θ_{sat} can be treated as an optimized parameter in the Rossi-Nimmo (1994) junction model, however, its value was known by direct measurement, and it was treated as a fixed parameter. ψ_d also was treated as a fixed parameter because its value can be calculated from the Kelvin equation, using average laboratory temperature during oven drying of samples and the relative humidity of water at average oven and room temperatures (Ross and others, 1991; Andraski, 1996). Under laboratory conditions of 50 percent relative humidity and 22°C, and an oven temperature of 105°C, ψ_d has a value near -8×10^6 cm-water. For convenience, a ψ_d value of -1×10^7 cm-water was used in the model fits for all core samples.

Particle-Size Parameters

To apply multiple linear-regression techniques, discrete measurements of the explanatory variables used to predict the hydraulic parameters (θ_{sat}, ψ_o, λ, and K_{sat}) are needed. These explanatory variables include ρ_{bulk}, ρ_{part}, and a representation of particle-size data. Because particle-size measurements spanned different size ranges and varied in resolution between data sets, the first step in representing particle-size data involved calculating cumulative particle-size distributions on a percent-finer-than basis. The cumulative particle-size distribution, like the $\theta(\psi)$ curve, is nonlinear. Three alternative representations of the particle-size distribution were chosen for comparison in developing the PTMs (fig. 6): (1) textural classes (percent gravel, sand, silt, and clay), (2) geometric statistics (geometric mean particle diameter (M_g) and geometric standard deviation (σ_g)), and (3) graphical statistics (median particle diameter (d_{50}) and uniformity coefficient (C_u)).
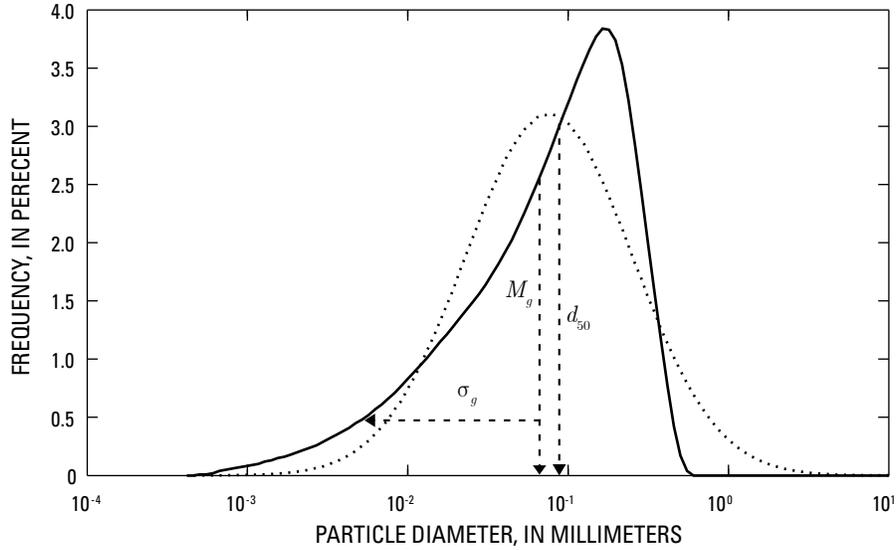
A. Differential frequency distribution showing the frequency of particles at a given particle diameter.
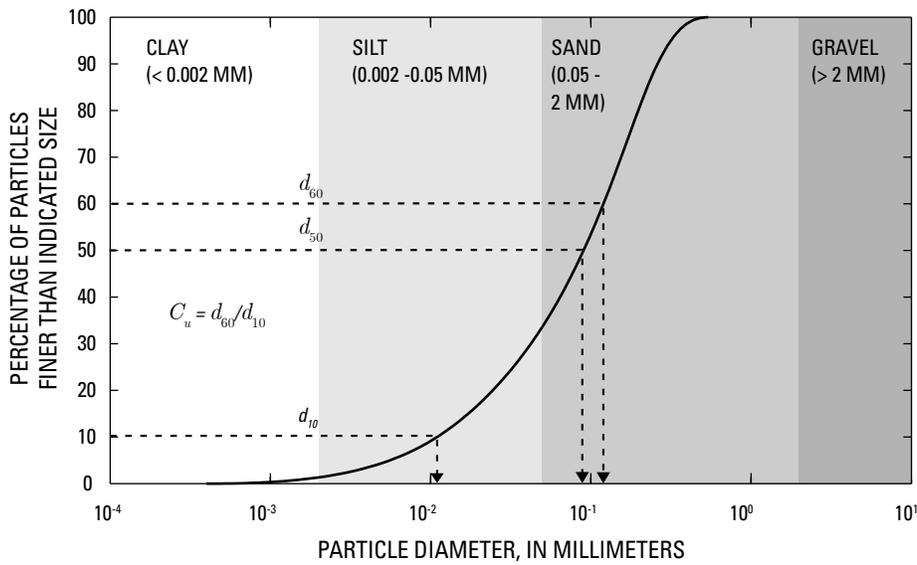
B. Cumulative particle-size distribution showing the frequency of particles finer than a given particle diameter.

EXPLANATION

M_g Geometric mean particle diameter

s_g Geometric standard deviation

d_{10} Diameter at the 10th percentile

d_{50} Median particle diameter

d_{60} Diameter at the 60th percentile

C_u Uniformity coefficient

— Measured data

..... Normal distribution

Shaded regions indicate particle-size limits defining the USDA textural classes (gravel, sand, silt, and clay).

Figure 6.   Example particle-size distribution showing three parameter sets (textural-class percentages, geometric statistics, and graphical statistics) used to develop property-transfer models for water retention and saturated hydraulic conductivity.

Textural-class percentages commonly are recorded in lithologic logs or are easily measured in the field or laboratory, requiring only a few measurements of particle size. Textural-class percentages also have been used to develop simple PTMs for other sites or applications (Gupta and Larson, 1979; Cosby and others, 1984; Puckett and others, 1985; Wösten and van Genuchten, 1988; Schaap and others, 1998, 2001; Schaap, 1999). For this reason, textural-class percentages were chosen as one way of representing the particle-size data. Weight percentages were calculated from the cumulative particle-size distributions using the definitions of gravel, sand, silt, and clay from the U.S. Department of Agriculture's (USDA) soil classification system (Soil Survey Staff, 1975; fig. 6B). Linear interpolation between points on the cumulative particle-size distribution was required to determine the frequency of particles falling between the particle-size limits defined by this soil classification system.

Because particle-size distributions of sediments typically follow a lognormal distribution (Krumbein, 1938; Pettijohn, 1975), descriptive statistical formulas, using the method of moments (Beyer, 1991, p. 470-471), were used to calculate M_g and σ_g. M_g is a summary measure of the most abundant particle diameter in a distribution. For a distribution with skewness, or excess particles in one-half the distribution compared to the other, M_g is weighted by the asymmetry away from the mean of the normal distribution in the direction of excess particles (fig. 6A). M_g and σ_g together define a particle-size distribution having a single mode. M_g was calculated by the formula:

$$\log_{10}\left(M_g\right) = \frac{\sum_{i=1}^{k}\left[f_i(d_{ci})\log_{10}(d_{ci})\right]}{\sum_{i=1}^{k}f_i(d_{ci})}, \quad (4)$$

where

k is the number of bins,

d_{ci} is the geometric center of the ith bin, and

$f_i(d_{ci})$ is the frequency of sizes occurring within ith bin assigned to d_{ci}.

σ_g was calculated by:

$$\left[\log_{10}\left(\sigma_g\right)\right]^2 = \frac{\sum_{i=1}^{k}\left\{f_i\left(d_{ci}\right)\left[\log_{10}\left(d_{ci}\right)-\log_{10}\left(M_g\right)\right]^2\right\}}{\sum_{i=1}^{k}f_i\left(d_{ci}\right)}. \quad (5)$$

The third representation of particle-size data uses graphical statistics, where particle diameters are determined at specific cumulative frequencies (fig. 6B). The midpoint of the cumulative particle-size distribution is the particle diameter at the 50th percentile, or d_{50}. One disadvantage of using d_{50}

is that it excludes the influence of distribution extremes, so skewed distributions are not well represented (Folk, 1980). A measure of the spread in particle sizes, referred to as the uniformity coefficient (C_u), is calculated as the diameter at the 60th percentile (d_{60}) divided by the diameter at the 10th percentile (d_{10}). When the cumulative particle-size distribution is plotted on a log-log scale, these percentiles commonly lie in the part of the distribution that can be fit with a straight line, and hence, C_u describes the slope (Fetter, 1994). A small C_u reflects a steep curve, or a narrow range of particle sizes. The diameter at any given percentile can be determined graphically by drawing a horizontal line from the percentile of interest, usually plotted on the y-axis, until it intersects the particle-size distribution. A vertical line, perpendicular to the horizontal line, is drawn until it meets the x-axis, where the particle diameter is then read. Because a large number of samples needed to be processed in this study, a computer program was used to calculate d_{10}, d_{50}, and d_{60} by linearly interpolating between the two points on the cumulative particle-size distribution nearest the percentile of interest.

Multiple Linear Regression

Multiple linear-regression analyses were done using custom programs written with Matlab (The MathWorks, Inc., Version 6, Release 12) and the Statistics Toolbox utility. The basic steps of the multiple linear-regression procedure used in this study are outlined in figure 4. Hydraulic property data included measurements of θ_{sat}, K_{sat}, and curve-fit parameters (ψ_o and λ) for the $\theta(\psi)$ measurements. Bulk-physical properties included ρ_{bulk}, ρ_{part}, and particle-size parameters (textural-class percentages; M_g and σ_g; d_{50} and C_u). Hydraulic properties (referred to as response variables) were predicted from a linear combination of the bulk-physical properties, referred to as explanatory variables.

The basic functional form for multiple linear regression is:

$$y_j = \sum_{i=0}^{p}b_i x_{ij} + e_j, \, j = 1,2,...n, \quad (6)$$

where

n is number of observations (note, in this application, n refers to the number of core samples with complete sets of measured hydraulic and bulk-physical properties),

y_j is observed value of the response variable for the jth core sample,

p is number of x (explanatory) variables including an intercept term,

b_i is a coefficient estimated by regression and associated with x_i, and

e_j is prediction error term, or residual.

To calculate the regression equations, a complete set of hydraulic and bulk-physical-property measurements was necessary for each core sample; otherwise, the sample was removed from the analysis. The right-hand side of the equation that includes regression coefficients and explanatory variables is referred to as the regression variate. An intercept term ($i = 0$) must be included by setting $x_0 = 1$ for each core sample. For each core sample, e_j is calculated by subtracting the estimated value of the response variable from the observed value. Positive residuals indicate that the response variable is underestimated by the regression variate. The least-squares procedure yields the best estimates for b_i by minimizing the objective function:

$$S(b) = \sum_{j=1}^{n} \left(y_j - \sum_{i=0}^{p} b_i x_{ij} \right)^2 . \tag{7}$$

The strength of the linear relation between a set of explanatory variables and a given response variable (goodness-of-fit) was assessed using the adjusted coefficient of determination (R^2_{adj}), which accounts for the number of observations and the number of explanatory variables. R^2_{adj} is directly related to the coefficient of determination (R^2), defined as:

$$R^2 = \frac{SS_{reg}}{SS_{res} + SS_{reg}} , \tag{8}$$

where

SS_{res} is the sum of squared residuals (equivalent to equation 7 using the known values of b_i) and

SS_{reg} is the regression sum of squares.

SS_{reg} is calculated by:

$$SS_{reg} = \sum_{j=1}^{n} \left(\sum_{i=0}^{p} b_i x_{ij} - \frac{\sum_{j=1}^{n} y_j}{n} \right)^2 , \tag{9}$$

Then R^2_{adj} is calculated by:

$$R^2_{adj} = 1 - \frac{(n-1)}{(n-p)}(1 - R^2) . \tag{10}$$

R^2_{adj} measures the ability of the explanatory variables in the regression variate to explain variation in the response variable. A value of R^2_{adj} equal to 1 indicates that the explanatory variables are able to explain all variation in the response variable. For comparing the ability of one regression variate (A) to explain variation in y over another regression variate (B), the percent change in R^2_{adj} is calculated by $100(R^2_{adj}{}^{(A)} - R^2_{adj}{}^{(B)})$.

The root-mean-square error (*RMSE*), also referred to as the standard error of the estimate, is a measure of the variation in the predicted values. It can be used to differentiate between competing models of a single response variable. *RMSE* (in units of the response variable) is calculated as:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{(n-p)}} , \tag{11}$$

where $\hat{y}_j$ is the predicted value of the response variable (equivalent to the second term in equation 7). Smaller values of *RMSE* indicate that the explanatory variables of a particular regression variate are more successful in explaining y than those in competing models.

Statistical tests are used to make inferences about how representative the results of a regression analysis, based on a particular sample, are of the population. In this context, sample refers to the group of observations (hydraulic and bulk-physical properties). Significance testing involves stating a hypothesis, specifying a significance level (α) for the test (probability of finding a hypothesis true when, in fact, it is not), and determining whether the hypothesis is true or false. The hypothesis is rejected if the value of the test statistic exceeds a certain critical value, calculated using the α-level, number of observations, and number of explanatory variables. In this study, the significance level (α) was specified at 0.05. For the significance tests to be valid and to yield accurate results, the assumption of normally-distributed residuals must be met.

For the t-test, the hypothesis states that the value of a particular regression coefficient is equal to zero. Partial t-values were calculated by dividing the value of the coefficient by its standard error. The standard error denotes the expected variation of a coefficient across multiple samples of data from a given population. Smaller standard error values imply more reliable predictions (smaller confidence intervals for the coefficient). To determine whether a regression coefficient is significantly different from zero, the partial t-value is compared to the critical t-value, which is determined from the Student's t-distribution for the number of observations and α-level specified. If the magnitude of the partial t-value exceeds the critical value, then the coefficient is significantly different from zero, indicating that the explanatory variable is useful in explaining the variation in the response variable.

For purposes of predicting the response variable from a set of explanatory variables, hypothesis testing, and estimating confidence intervals, the assumptions of multiple linear regression are that

1. Residuals are normally distributed,

2. Response variable displays equal levels of variance across the range of values for the explanatory variables (homoscedasticity),

3. Relation between the response variable and each explanatory variable is linear, and

4. No correlated errors or systematic effects exist in the data.

The assumption that the residuals are normally distributed must be met for valid hypothesis testing or calculation of confidence intervals.

   To determine whether the residuals were normally distributed, the residuals were plotted on a normal probability plot and the probability-plot correlation coefficient (PPCC) was calculated. PPCC measures the departure of the residual distribution from a straight line, which represents the normal distribution on a probability plot. A normal distribution would have a PPCC value of 1.0 and skewness of zero. Skewness was used as another means of assessing normality. Skewness measures the degree of asymmetry in a distribution, with negative skewness indicating a distribution with an excess of smaller values compared to the mean of the normal distribution. Skewness should be near zero for a residual distribution to be considered normal, and the PPCC value should be near or greater than the critical value for the number of observations comprising the distribution (Looney and Gulledge, 1985; Helsel and Hirsch, 1992).

   Residual plots were used to identify heteroscedastic trends (unequal variance in either the response variable or one of the explanatory variables) that would violate a basic assumption of multiple linear regression. A residual plot is constructed by plotting the residuals from a particular regression model against the predicted values of the response variable. More commonly the residual is studentized. This refers to a special form of standardization where the jth observation is omitted. Standardization is used to create a common scale for comparing the residuals from different models. Each residual is standardized by dividing its value by the standard deviation of residuals for all observations from a particular regression analysis. After standardization, the residual distribution has a mean value of 0 and a standard deviation of 1 (Hair and others, 1998). For n greater than or equal to 50, the studentized residual distribution approximates the Student's t-distribution. Residuals whose absolute values exceed the critical t-value (typically near 2) are considered outliers. Heteroscedasticity exists if the pattern of residuals appears nonrandom, with either a nonlinear trend, clustering of points in a particular region, or both. Heteroscedasticity can be corrected, although sometimes not entirely, by transforming one or more explanatory variables or the response variable.

   Linearity between response and explanatory variables was assessed by examining scatter plots and variable distributions, and by calculating Pearson's correlation coefficient (r). Scatter plots of each response variable and each explanatory variable, in addition to Pearson's r, were used to identify univariate trends in the data. Pearson's r measures the correlation between two variables. It can vary from -1 to +1, with 0 indicating no correlation and ± 1 indicating perfect correlation.

   While Pearson's r is used to identify univariate trends in the data and to identify explanatory variables that exhibit high covariance, the variance inflation factor (VIF) is used to test for multicollinearity resulting from the combined effect of two or more variables in the regression variate. VIF indicates the effect of the other explanatory variables on the explanatory variable of interest; each explanatory variable (x_i) is in turn regressed against the remaining explanatory variables to calculate R_i^2 (as defined in equation 8), where $VIF = 1/(1-R_i^2)$. Large values of VIF, typically greater than 10, indicate a high degree of collinearity or multicollinearity among the explanatory variables examined (Hair and others, 1998).

   The existence of collinearity between explanatory variables invalidates hypothesis tests, for example if interpretation of estimated coefficients (b_i) is important. However, if prediction is the main application of the regression models, then collinearity is acceptable as long as the prediction is made within the range of data values used in model development (Helsel and Hirsch, 1992). Multicollinearity makes it difficult to separate the effects among explanatory variables on the regression model, for example when stepwise procedures are used to iteratively select variables. Other effects of multicollinearity include a limited size to R^2 (because the predictive power of each explanatory variable is shared or overlaps with other variables), incorrect magnitudes or signs of the coefficients, and instability in the coefficient values during the regression (Helsel and Hirsch, 1992; Hair and others, 1998).

   Another assumption of multiple linear regression is that the predicted values of the response variable are not related to each other, or that the values are not sequenced by any particular explanatory variable. Sequencing often occurs when time- or event-based variables are included in the regression analysis. Plotting residuals against a given explanatory variable should show a random pattern if the predicted values are independent. A clustered or nonlinear pattern in this type of plot indicates error dependence, or that an essential explanatory variable has been omitted from the regression variate (Hair and others, 1998). Hydraulic and bulk-physical properties of core samples used to develop multiple linear-regression equations in this study should be independent of seasonal fluctuations in moisture content and timing of core collection, therefore, the error terms for each regression model were considered independent without further testing of this assumption.

Property-Transfer Model Analyses

   Multiple linear-regression techniques were used to develop two separate PTMs, one for $\theta(\psi)$ and the other for K_{sat}. The $\theta(\psi)$ PTM involved a separate regression equation for each of the three parameters (θ_{sat}, ψ_o, and λ) that define the Rossi-Nimmo (1994) junction model. For each of the four

hydraulic properties (θ_{sat}, ψ_o, λ, and K_{sat}), three models were examined to identify the best set of particle-size parameters to use as explanatory variables. Model A refers to the regression variate that includes ρ_{bulk}, ρ_{part}, and textural-class percentages (gravel, sand, silt, and clay). Model B includes ρ_{bulk}, ρ_{part}, M_g, and σ_g as explanatory variables. Model C includes ρ_{bulk}, ρ_{part}, d_{50}, and C_u.

The minimum, maximum, average, and standard deviation of the hydraulic properties (θ_{sat}, ψ_o, λ, and K_{sat}), (table 2), and bulk-physical properties (ρ_{bulk}, ρ_{part}, and various particle-size parameters), (table 3), are summarized for each data source. Although ψ is a negative pressure measured relative to atmospheric pressure, absolute ψ values were used to develop the regression equations for ψ_o. Textural classes (gravel, sand, silt, and clay) are reported as weight percentages, or the weight of material falling within a given range of particle diameters divided by the total weight of the material analyzed. Particle-size limits for each textural class are defined according to the U.S. Department of Agriculture's soil classification system (Soil Survey Staff, 1975). Ninety-one out of 108 samples were classified as sands or loams, with 38 samples classified as silt loams. The texture of sample USGS 118 at 173.48 to 174.39 m from McElroy and Hubbell (1990) could not be determined due to incomplete particle-size data. The Magnuson (written commun., 2002) data set included the finest-textured samples, with 17 samples classified as clay loams, silty clay loams, silty clays, and clays. Because gravel-sized particles (> 2 mm) were removed prior to measuring $\theta(\psi)$ of the Magnuson (written commun., 2002) samples, the particle-size distributions of this data set were adjusted to zero percent gravel before calculating the particle-size parameters used to develop the $\theta(\psi)$ PTMs. In table 3, two summaries of the particle-size parameters are reported for Magnuson (written commun., 2002), one for use in developing the K_{sat} PTM and the other for developing the $\theta(\psi)$ PTM.

Evaluation of Calibration Data

During the multiple linear-regression analyses, data were thoroughly examined for typographical errors, measurement errors, or differences between groups of core samples (arising from differences in measurement technique, collection method, or sample handling). In some cases, core samples were removed from the regression analyses due to missing or incomplete data.

Effect of Repacked Core Samples

For the Magnuson (written commun., 2002) data set, $\theta(\psi)$ measurements were completed on disturbed samples (after removing gravel and repacking). For the other data sets used in this study, $\theta(\psi)$ curves were measured on intact samples. Although θ_{sat} was not measured on the Magnuson (written

commun., 2002) samples, the repacking effect was observed by comparing the total porosity of the repacked samples (taken as the θ_{sat} values for these samples) with the total porosity calculated from the intact samples used for measurement of K_{sat}. Measured θ_{sat} values should be less than or equal to the values of total porosity calculated for intact core samples. For the Magnuson (written commun., 2002) samples, total porosity of the repacked samples exceeded total porosity of the intact samples by 25 percent or more for 37 out of 51 samples. In contrast, only 3 out of 58 of the intact samples from the other data sets had measured θ_{sat} values that exceeded total porosity by 25 percent. Because the interbed samples are deep and consolidated, repacking "fluffs" the material, creating a larger mean pore size compared to an intact sample. As a result, the repacked ρ_{bulk} will be smaller than the ρ_{bulk} of the intact core sample. The repacked total porosity is then larger than that of the intact sample.

"Fluffing" the material by repacking also should affect ψ_o and λ values because the sample mean pore size has been increased. From capillary theory, ψ is inversely related to the pore size at which drainage occurs. An increase in mean pore size as a result of repacking suggests that the absolute value of ψ_o would be smaller than for an identical, intact sample, with drainage occurring "earlier" on the $\theta(\psi)$ curve. A more uniform pore-size distribution is expected for a repacked sample because the material has been homogenized, destroying both large- and small-sized pores in the process. This would be reflected in the $\theta(\psi)$ curve by a steeper drainage slope or a larger λ value.

The Mann-Whitney rank-sum test (Zar, 1996) was used to determine if the medians of the distributions of θ_{sat}, ψ_o, and λ for repacked (S. Magnuson, written commun., 2002) and intact (McElroy and Hubbell, 1990; Perkins and Nimmo, 2000; Perkins, 2003; Winfield, 2003) samples were equal. For θ_{sat}, the calculated test statistic, Z_c, was 3.191, compared to the critical two-tailed value of 1.645, using the t-distribution approximation to a normal distribution with infinite degrees of freedom and an α-level of 0.10. For ψ_o, Z_c was 2.930, and for λ, Z_c was 1.473. The rank-sum test indicated a statistical difference between repacked and intact samples for θ_{sat} and ψ_o. Aside from the way samples were pretreated before experimental analyses, this difference could be due to the high clay content of the Magnuson samples (table 3) compared to the other data sets. However, the test was repeated after excluding 17 samples classified as clay loams, silty clay loams, silty clays, and clays from the Magnuson data set. Results showed that the medians of the θ_{sat} and ψ_o distributions between intact and repacked samples were again dissimilar, suggesting that repacking has a stronger effect on the distributions of θ_{sat} and ψ_o than percent clay does. For λ, the distributions of repacked and intact samples were found to be dissimilar after excluding the samples with high clay content, indicating that repacking also affects λ.

Table 2. Statistics for hydraulic data used to develop property-transfer models, listed by data source.

[**Abbreviations**: cm^3/cm^3, cubic centimeter per cubic centimeter; cm-water, centimeter of water; cm/s, centimeter per second]

Statistic	Saturated water content, θ_{sat} (cm^3/cm^3)	Scaling parameter for matric pressure, ψ_o (-cm-water)	Curve-shape parameter, λ	Saturated hydraulic conductivity, K_{sat} (cm/s)
Magnuson, written commun. (2002)[1]				
Minimum	0.4427	0.06	0.0820	1.14×10^{-7}
Maximum	.7333	241.77	.6568	8.41×10^{-2}
Average	.5098	45.73	.2380	4.03×10^{-3}
Standard deviation	.0482	54.63	.1292	1.32×10^{-2}
Number of samples	51	51	51	41
McElroy and Hubbell (1990)				
Minimum	0.3481	7.97	0.0853	1.13×10^{-8}
Maximum	.6290	3,499.98	.7082	6.73×10^{-3}
Average	.5191	295.56	.2340	1.33×10^{-3}
Standard deviation	.0814	831.84	.1836	2.20×10^{-3}
Number of samples	16	16	16	16
Perkins and Nimmo (2000)				
Minimum	0.3473	3.13	0.1066	2.35×10^{-7}
Maximum	.4497	556.67	1.1850	3.90×10^{-3}
Average	.4083	150.55	.3056	8.73×10^{-4}
Standard deviation	.0326	179.37	.2607	1.62×10^{-3}
Number of samples	18	18	18	18
Perkins (2003)				
Minimum	0.4516	2.12	0.0765	1.66×10^{-7}
Maximum	.5773	399.93	.5312	1.42×10^{-3}
Average	.5075	67.79	.1742	2.78×10^{-4}
Standard deviation	.0371	106.36	.1249	3.81×10^{-3}
Number of samples	14	14	14	12
Winfield (2003)				
Minimum	0.3585	16.65	0.0950	2.07×10^{-7}
Maximum	.5661	1,207.78	.2593	8.31×10^{-4}
Average	.4796	430.15	.1681	8.88×10^{-5}
Standard deviation	.0531	334.98	.0543	5.34×10^{-3}
Number of samples	10	10	10	10
Total number of samples, n	109	109	109	97

[1]Water-retention measurements were completed on samples that were repacked after removing gravel (particle sizes greater than 2 millimeters).

Table 3.   Statistics for bulk physical-property data used to develop property-transfer models, listed by data source.

[The uniformity coefficient is defined as the particle diameter at the 60th percentile divided by the diameter at the 10th percentile, calculated from the cumulative particle-size distribution on a percent-finer-than basis. **Abbreviations:** g/cm^3, gram per cubic centimeter; mm, millimeter]

Statistic	Bulk density, ρ_{bulk} (g/cm^3)	Particle density, ρ_{part} (g/cm^3)	Gravel (percent)	Sand (percent)	Silt (percent)	Clay (percent)	Geometric mean particle diameter, M_g (mm)	Geometric particle-size standard deviation, σ_g	Median particle diameter, d_{50} (mm)	Uniformity coefficient, C_u
Magnuson, S., written commun. (2002)[1]										
Minimum	1.06	2.04	0.00	0.89	3.33	1.76	0.0013	2.76	0.0009	2.94
Maximum	1.97	2.66	74.70	91.05	70.07	63.81	2.4070	71.43	5.5197	2,254.05
Average	1.58	2.42	11.79	34.77	33.23	20.20	.2280	10.70	.4035	85.85
Standard deviation	.21	.13	18.04	25.14	18.95	18.35	.5154	11.16	1.1436	311.74
Number of samples	51	51	51	51	51	51	51	51	51	51
Magnuson, S., written commun. (2002), adjusted[2]										
Minimum	1.06	2.04	0.00	0.89	3.92	2.17	0.0012	0.00	0.0007	2.86
Maximum	1.97	2.66	.00	92.92	70.07	81.40	.2427	14.76	.6163	257.62
Average	1.58	2.42	.00	41.92	36.25	21.83	.0568	6.32	.0825	30.78
Standard Deviation	.21	.13	.00	29.25	17.87	19.18	.0678	2.84	.1216	42.50
Number of samples	51	51	51	51	51	51	51	51	51	51
McElroy and Hubbell (1990)										
Minimum	1.18	2.52	0.00	21.83	2.84	0.12	0.0133	1.91	0.0115	2.59
Maximum	1.94	2.89	.26	96.91	67.36	15.75	.2363	7.78	.2345	79.44
Average	1.44	2.63	.03	61.37	32.12	6.50	.0708	4.27	.0820	20.45
Standard deviation	.20	.08	.07	24.11	21.10	5.08	.0720	1.59	.0699	24.01
Number of samples	16	16	15	15	15	15	15	15	15	15
Perkins and Nimmo (2000)										
Minimum	1.22	2.65	0.00	0.04	2.37	0.14	0.0065	2.14	0.0082	2.77
Maximum	1.63	2.65	6.81	97.25	82.88	17.83	.4267	8.02	.4395	32.95
Average	1.42	[3]2.65	.82	34.73	53.31	9.97	.0881	4.22	.0995	13.61
Standard deviation	.09	.00	1.92	33.03	29.62	6.04	.1328	1.37	.1437	7.34
Number of samples	18	18	18	18	18	18	18	18	18	18
Perkins (2003)										
Minimum	1.11	2.60	0.00	11.86	7.76	1.29	0.0094	2.71	0.0108	3.30
Maximum	1.46	2.75	.02	90.97	74.13	13.99	.1992	9.42	.3970	118.17
Average	1.31	2.63	.00	59.20	34.54	6.28	.0856	5.75	.1361	32.70
Standard deviation	.10	.04	.00	24.97	20.95	4.20	.0579	2.08	.1145	33.33
Number of samples	14	14	14	14	14	14	14	14	14	14
Winfield (2003)										
Minimum	1.18	2.74	0.00	0.00	61.43	9.91	0.0039	3.01	0.0046	6.98
Maximum	1.63	2.84	.00	25.20	80.80	25.08	.0160	4.67	.0183	16.40
Average	1.41	2.78	.00	9.00	73.59	17.41	.0083	3.87	.0097	10.95
Standard deviation	.12	.03	.00	8.85	5.45	4.44	.0036	.55	.0042	2.83
Number of samples	10	10	10	10	10	10	10	10	10	10
Total number of samples, n	109	109	108	108	108	108	108	108	108	108

[1] Core sample properties were used to develop the saturated hydraulic conductivity property-transfer model.

[2] Adjusted samples: water-retention measurements were completed on samples that were repacked after removing gravel (particle sizes greater than 2 mm). Particle-size distributions were normalized to zero gravel content prior to calculating parameters. These core sample properties were used to develop the water-retention property-transfer model.

[3] The average particle density of 2.65 g/cm^3, for measurements on 4 core samples, was used for all 18 samples.

Including the repacked samples in preliminary regression analyses had a pronounced effect on the results compared to analyses that excluded these samples. The goodness-of-fit values, indicated by R^2_{adj}, decreased by 30.9 to 31.2 percent for models A, B, and C for θ_{sat}, by 22.8 to 39.6 percent for ψ_o, and by 4.2 to 34 percent for λ. Note that for the Magnuson data set, M_g, σ_g, d_{50}, and C_u were determined on particle-size distributions reconstituted without gravel (to match the way the $\theta(\psi)$ measurements were completed), although ρ_{bulk} values included gravel. To adjust the measured ρ_{bulk} values to zero percent gravel, the ρ_{bulk} of the gravel fraction would have to be known (Andraski, 1996). As a result of these findings, the Magnuson data were excluded from the development of the $\theta(\psi)$ PTMs. However, this data set was included in the development of the K_{sat} PTM because K_{sat} and ρ_{bulk} were measured on intact core samples.

Errors in Fitted, Calculated, or Measured Parameters

Complete core sample measurement sets were needed to do the regression analyses for θ_{sat}, ψ_o, λ, and K_{sat}. Because some core samples had missing values, the number of core samples available for developing the PTMs was reduced. One sample from McElroy and Hubbell (1990), USGS 118 at 173.48 to 174.39 m, had only partial particle-size data. Because the particle-size parameters for this sample could not be calculated, this sample was excluded from all regression analyses, reducing the number of samples for developing the $\theta(\psi)$ PTMs from 109 to 108 (table 4). K_{sat} measurements were not completed for 10 samples from Magnuson (written commun., 2002) and two samples from Perkins (2003). The total number of samples for developing the K_{sat} PTM was reduced from 109 to 96, accounting for the sample with missing particle-size data and the 12 samples with missing K_{sat} values (table 4).

Table 4.   Core samples removed from multiple linear-regression analyses due to missing data.

[Model A, regression model with textural class percentages and bulk density as explanatory variables; Model B, regression model with geometric mean particle diameter, geometric standard deviation, and bulk density as explanatory variables; Model C, regression model with median particle diameter, uniformity coefficient, and bulk density as explanatory variables. **Abbreviations:** m, meter; B, Core sample removed due to error in one or more bulk physical-properties, typically a particle-size parameter; H, Core sample removed due to error in the hydraulic property value indicated]

| Borehole identification No. | Depth interval (m) | Water retention ($\theta(\psi)$) parameters | | | | | | | | | Saturated hydraulic conductivity, $\log(K_{sat})$ | | |
| | | Saturated water content, θ_{sat} | | | Scaling parameter for matric pressure, $\log(\psi_o)$ | | | Curve-shape parameter, $\log(\lambda)$ | | | | | |
		Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C
Magnuson, S., written commun. (2002)													
RWMC-SCI-V-153	32.74–33.54										H	H	H
RWMC-SCI-V-154	33.99–34.30										H	H	H
RWMC-SCI-V-157	71.04–71.19										H	H	H
RWMC-SCI-V-159	72.56–72.71										H	H	H
RWMC-SCI-V-160	73.84–73.99										H	H	H
SOUTH-SCI-V-012	35.98–36.13										H	H	H
SOUTH-SCI-V-013	71.98–72.13										H	H	H
SOUTH-SCI-V-014	35.18–35.37										H	H	H
	36.40–36.59										H	H	H
SOUTH-SCI-V-018	74.21–74.39										H	H	H
McElroy and Hubbell (1990)													
USGS 118	173.48–174.39	B	B	B	B	B	B	B	B	B	B	B	B
Perkins (2003)													
ICPP-SCI-V-215	45.53–45.59										H	H	H
	45.59–45.65										H	H	H
Total number of samples		109	109	109	109	109	109	109	109	109	109	109	109
Number of samples with missing data		1	1	1	1	1	1	1	1	1	13	13	13
Adjusted total number of samples, n		108	108	108	108	108	108	108	108	108	96	96	96

Because errors in hydraulic and bulk-physical properties can cause the regression models to be strongly influenced by a few points, the values of the explanatory and response variables were carefully examined. Errors can be due to sample collection, handling, measurement technique, and measurement resolution. Errors in sample collection and handling can adversely affect the ρ_{bulk} of a core sample, which needs to be representative of a minimally disturbed sample. Sample disturbance also affects the hydraulic properties, particularly those most sensitive to the structure of the medium, such as θ_{sat} or K_{sat}. The resolution of the $\theta(\psi)$ measurements can affect the quality of the Rossi-Nimmo (1994) junction curve fits and cause errors in the optimized curve-fit parameters, ψ_o and λ. The textural, geometric, and graphical parameters calculated from particle-size distributions also are sensitive to the resolution and range of measurements.

Measurements of ρ_{bulk} and ρ_{part} were examined for errors by dividing the values into each of the nine USDA textures represented by the data (fig. 7A). ρ_{bulk} is a complex property affected by multiple factors, such as depth (or amount of overburden pressure), texture, particle arrangement, and aggregation or macropores. The influences of these factors, in addition to the possible effects of sample disturbance during core collection and subsequent handling, on ρ_{bulk} are difficult to separate. Because natural variation in ρ_{bulk} is expected, no values of ρ_{bulk} seemed unusually high or low for a given texture. Values of ρ_{part} ranged from 2.04 to 2.89 g/cm^3 (table 3). Minerals commonly found in the sedimentary interbeds beneath the RWMC include quartz, plagioclase, orthoclase, pyroxene, and calcite (Rightmire and Lewis, 1987; Bartholomay, 1990; Reed and Bartholomay, 1994). The ρ_{part} values of these minerals range from 2.57 to 2.76 g/cm^3. Compared to these values, the ρ_{part} values of the other data sets, and the ρ_{part} values of similarly textured samples, one value of ρ_{part} equal to 2.04 g/cm^3 (for sample SOUTH-SCI-V-015 at 38.81 to 39.02 m, classified as a clay) seemed too low; therefore, this sample was excluded from the regression analyses.

Several samples were excluded from the $\theta(\psi)$ and K_{sat} PTMs because their particle-size distributions did not extend to small enough diameters (or cumulative frequencies) to accurately calculate statistical and textural-class parameters. The resolution of the particle-size distribution also will affect the values of most particle-size parameters; however, this issue is not addressed in this report. To adequately define the distribution shape, it is important to have as many points at the particle-size distribution's fine end as at the coarse end. Definition at the distribution's fine end is needed to interpolate d_{10} (for calculating C_u) and the frequency at 0.002 mm (for determining percent clay and percent silt). If the cumulative frequency at the smallest measured particle size is greater than 10 percent, then d_{10} can have significant error because this part of the cumulative distribution's slope is unknown.
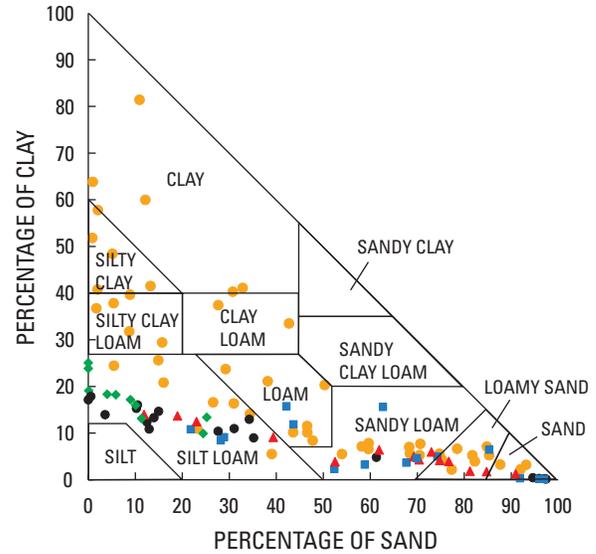
A. All potential core samples, except those with missing data (n = 108).
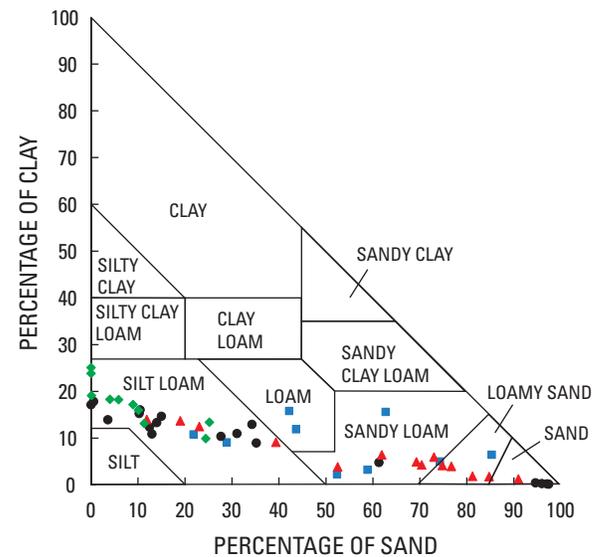
B. Core samples used in model development, shown with the maximum number removed due to errors and missing data (n = 51).

EXPLANATION

DATA SOURCE

- ● RWMC: Magnuson (2002, written commun.)
- ■ RWMC: McElroy and Hubbell (1990)
- ● RWMC: Perkins and Nimmo (2000)
- ▲ VZRP: Perkins (2003)
- ◆ VZRP: Winfield (2003)

RWMC, Radioactive Waste Management Complex

VZRP, Vadose Zone Research Park

Figure 7. Textural classification of core samples (based on the system of the U.S. Department of Agriculture) used to develop a property-transfer model for estimating water retention ($\theta(\psi)$).

If a significant percentage of particles is finer than the smallest measured particle diameter, then the width of the last bin at the distribution's fine end, defined as the interval between the smallest measured diameter and zero, will be large, weighting M_g toward larger values and σ_g toward smaller values. The cumulative frequency at which deviations in M_g and σ_g became unacceptable was near the 10th percentile. Therefore, samples that did not have measurements on the cumulative particle-size distribution that extended to at least the 10th percentile were excluded from the multiple linear-regression analyses. Samples that did not have particle-size measurements extending to at least 0.002 mm were excluded because percent clay and percent silt could not be determined. Four McElroy and Hubbell (1990) samples (USGS 118 at 173.48 to 174.39 m, D-02 at 70.91 to 71.41 m, D-02 at 71.41 to 71.71 m, and D-15 at 33.11 to 33.84 m) were excluded from model A due to lack of particle-size data in the clay range (down to 0.002 mm). For models B and C, three samples from this data set (USGS 118 at 173.48 to 174.39 m, D-15 at 33.11 to 33.84 m, and D-15 at 69.94 to 71.16 m) were excluded because the size of the last particle-size bin was too large, causing errors in the calculated values of M_g, σ_g, and C_u. For the Magnuson data set, all particle-size measurements extended to small enough diameters to allow determination of percent silt and percent clay. However, 26 out of 51 samples were excluded from models B and C ([table 5](#)) due to poor definition at the fine end of the particle-size distribution, which affected the values of M_g, σ_g, and C_u. These 26 samples were classified as silt loams, clay loams, silty clay loams, silty clays, and clays. Among the K_{sat} models, these samples are the finest in texture compared to the remaining samples ([fig. 8A](#)); as a result, model A included more fine-textured samples than models B and C.

Values of each hydraulic property (θ_{sat}, ψ_o, λ, and K_{sat}) were evaluated by examining their distributions within each of the nine USDA textures. A few anomalous values of K_{sat} and ψ_o were identified this way. One core sample, SOUTH-SCI-V-011 at 32.29 to 32.44 m, from the Magnuson data set had a measured K_{sat} value of 1.24×10^{-3} cm/s, which is too large for a sample containing 41 percent clay. This sample was removed from the final regression analyses for K_{sat}. Cracks in the sample or gaps between the sample and sample holder could have contributed to this large K_{sat} measurement. Swelling of a sample containing a large amount of clay would tend to lower K_{sat}. The maximum ψ_o value of all 58 samples used for calibrating the $\theta(\psi)$ models (sample D-15 at 33.11 to 33.84 m from the McElroy and Hubbell (1990) data set) was 3,499 cm for a sandy sample. Because this ψ_o value is much too large for a sample of this texture, sample D-15 at 33.11 to 33.84 m was removed from the ψ_o regression models. Values of ψ_o are affected by the texture (particle-size distribution) and structure

(particle arrangement, macropores) of the porous media, the influences of which are difficult to separate. For fine-textured samples and samples with a wide range of particle sizes, ψ_o will tend to be large. For sample D-15 at 33.11 to 33.84 m, the optimization of ψ_o may have been weighted by a few erroneous points in the wet range of θ. A mismatch of reported particle-size distribution with measured $\theta(\psi)$ curve (McElroy and Hubbell (1990) listed the texture of this sample as 'clay') also could be the reason for this large ψ_o value.

To evaluate the θ_{sat} measurements, values for each core sample were compared to total porosity values calculated from ρ_{bulk} and ρ_{part}. Values of θ_{sat} were converted to percent saturation by dividing by total porosity and multiplying by 100. Values of percent saturation exceeding 125 percent were removed from the regression analyses for θ_{sat}. This conservative threshold was selected to keep most samples in the analyses, but exclude samples with unreasonably high values of saturation compared to total porosity. Three core samples from the McElroy and Hubbell (1990) data set (from boreholes 88-01D, D-10, and USGS 118) had θ_{sat} values that met this criterion. Because the ρ_{bulk} values seemed reasonable for these three samples, ranging from 1.29 to 1.57 g/cm^3, the high θ_{sat} values could be due to sample disturbance (although the K_{sat} values of these samples are reasonable) or to saturation in excess of total porosity. As mentioned above, θ_{sat} was not measured for the Magnuson (written commun., 2002) core samples. For this reason, and due to repacking effects on the Rossi-Nimmo (1994) junction model parameters, all 51 of the Magnuson (written commun., 2002) samples were excluded from the $\theta(\psi)$ PTMs.

The λ parameter approximates the slope of the drainage portion of the $\theta(\psi)$ curve on a semilog scale, with larger values typically associated with coarse-textured materials or materials with a narrow range of particle sizes. The maximum value of λ was 1.19 for a sandy sample (UZ98-2 at 43.09 to 43.21 m from Perkins and Nimmo (2000)) and the minimum was 0.077 for a sandy loam sample (ICPP-SCI-V-215 at 46.10 to 46.20 m from Perkins (2003)). Because no upper limit on λ exists, there is no reason to omit the 1.19 value from the regression analyses. The minimum value is low for a sandy loam material but this could be due to the way the particles are arranged in the sample, or the particular proportions of sand, silt, and clay. The values of λ and ψ_o are affected by the number and quality of measurements defining the $\theta(\psi)$ curve. For sample ICPP-SCI-V-215 at 46.10 to 46.20 m, most of the measured points are clustered between 20 and 400 cm-water. Lack of $\theta(\psi)$ points in the dry range makes it difficult for the Rossi-Nimmo (1994) junction model to find an inflection point that would cause the curve fit to have a steeper drainage slope (corresponding to a larger λ value). However, no core samples were excluded from the multiple linear-regression analyses for having anomalous λ values.

Table 5. Core samples removed from multiple linear-regression analyses due to errors in fitted, calculated, or measured parameters.

[Model A, regression model with textural class percentages and bulk density as explanatory variables; Model B, regression model with geometric mean particle diameter, geometric standard deviation, and bulk density as explanatory variables; Model C, regression model with median particle diameter, uniformity coefficient, and bulk density as explanatory variables. **Abbreviations:** m, meter; B, core sample removed due to error in one or more of its bulk physical-properties, typically a particle-size parameter; H, core sample removed due to error in the hydraulic property value indicated]

Borehole identification no.	Depth interval (m)	Water retention ($\theta(\psi)$) parameters									Saturated hydraulic conductivity, $\log(K_{sat})$		
		Saturated water content, θ_{sat}			Scaling parameter for matric pressure, $\log(\psi_o)$			Curve-shape parameter, $\log(\lambda)$					
		Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C
Magnuson, written commun. (2002)													
RWMC-SCI-V-153	30.79–31.59	H	H	H	H	H	H	H	H	H			
	31.98–32.16	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	32.74–33.54	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
RWMC-SCI-V-154	30.34–31.55	H	H	H	H	H	H	H	H	H			
	32.16–32.32	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	33.99–34.30	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
RWMC-SCI-V-155	67.99–68.14	H	H	H	H	H	H	H	H	H			
RWMC-SCI-V-156	30.95–31.25	H	H	H	H	H	H	H	H	H			
RWMC-SCI-V-157	69.51–69.66	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	71.04–71.19	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
RWMC-SCI-V-158	29.73–29.94	H	H	H	H	H	H	H	H	H			
RWMC-SCI-V-159	68.17–68.35	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	69.82–70.27	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	72.56–72.71	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
	72.71–72.87	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
RWMC-SCI-V-160	68.29–68.45	H	H	H	H	H	H	H	H	H			
	72.29–72.74	H	H	H	H	H	H	H	H	H			
	73.84–73.99	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
SOUTH-MON-A-009	40.64–40.82	H	H	H	H	H	H	H	H	H			
	43.29–43.75	H	H	H	H	H	H	H	H	H			
	46.01–46.19	H	H	H	H	H	H	H	H	H			
SOUTH-SCI-V-011	29.88–31.89	H	H	H	H	H	H	H	H	H			
	32.29–32.44	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H
SOUTH-SCI-V-011	33.54–34.15	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	71.04–71.22	H	H	H	H	H	H	H	H	H			
SOUTH-SCI-V-012	33.54–33.96	H	H	H	H	H	H	H	H	H			
	35.98–36.13	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
	37.68–38.11	H	H	H	H	H	H	H	H	H			
SOUTH-SCI-V-013	29.45–29.88	H	H	H	H	H	H	H	H	H			
	69.70–71.55	H	H	H	H	H	H	H	H	H			
	71.04–71.22	H	H	H	H	H	H	H	H	H			
	71.98–72.13	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
SOUTH-SCI-V-014	31.40–31.65	H	H	H	H	H	H	H	H	H			
	35.18–35.37	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
	36.40–36.59	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
	73.78–73.96	H	H	H	H	H	H	H	H	H			
	74.21–74.39	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
SOUTH-SCI-V-015	35.67–35.88	H	H	H	H	H	H	H	H	H			
	36.89–37.10	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	38.81–39.02	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H	B	B	B
	70.73–70.95	H	H	H	H	H	H	H	H	H			
	75.09–75.30	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	77.53–77.74	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B

Table 5.   Core samples removed from multiple linear-regression analyses due to errors in fitted, calculated, or measured parameters.—Continued

[Model A, regression model with textural class percentages and bulk density as explanatory variables; Model B, regression model with geometric mean particle diameter, geometric standard deviation, and bulk density as explanatory variables; Model C, regression model with median particle diameter, uniformity coefficient, and bulk density as explanatory variables. **Abbreviations:** m, meter; B, core sample removed due to error in one or more of its bulk physical-properties, typically a particle-size parameter; H, core sample removed due to error in the hydraulic property value indicated]

| Borehole identification no. | Depth interval (m) | Water retention (θ(ψ)) parameters | | | | | | | | | Saturated hydraulic conductivity, $\log(K_{sat})$ | | |
| | | Saturated water content, θ_{sat} | | | Scaling parameter for matric pressure, $\log(\psi_o)$ | | | Curve-shape parameter, $\log(\lambda)$ | | | | | |
		Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C	Model A	Model B	Model C
Magnuson, written commun. (2002)—Continued													
SOUTH-SCI-V-016	37.96–38.11	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	41.55–41.77	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	45.21–45.43	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H		B	B
	74.70–74.85	H	H	H	H	H	H	H	H	H			
SOUTH-SCI-V-016	75.76–75.91	H	H	H	H	H	H	H	H	H			
SOUTH-SCI-V-018	68.69–68.87	H	H	H	H	H	H	H	H	H			
	71.04–71.25	H	H	H	H	H	H	H	H	H			
	74.21–74.39	H	B, H	B, H	H	B, H	B, H	H	B, H	B, H			
McElroy and Hubbell (1990)													
88-01D	[1]70.15–71.13	H	H	H									
D-02	70.91–71.41	B			B			B			B		
	71.41–71.71	B			B			B			B		
D-10	[1]10.58–10.98	H	H	H									
D-15	33.11–33.84	B	B	B	B, H	B, H	B, H	B	B	B	B	B	B
	69.94–71.16		B	B		B	B		B	B		B	B
USGS 118	68.60–70.12	H	H	H									
Perkins (2003)													
ICPP-SCI-V-215	45.53–45.59										H	H	H
	45.59–45.65										H	H	H
Total number of samples (excluding missing)		108	108	108	108	108	108	108	108	108	96	96	96
Number of samples removed due to errors		57	56	56	54	53	53	54	53	53	5	18	18
Adjusted total number of samples, *n*		51	52	52	54	55	55	54	55	55	91	78	78

[1]Two core samples were analyzed from the depth interval indicated. Samples indicated by McElroy and Hubbell (1990) as duplicates were removed.

Core samples that were removed from models A, B, and C for θ_{sat}, ψ_o, λ, and K_{sat} due to measurement errors, errors in fitted parameters, or errors in calculated parameters are listed in table 5. Samples removed due to errors in a measured or fitted hydraulic parameter are indicated with an 'H', while samples removed due to errors in a bulk property are indicated with a 'B'. The adjusted total number of samples reflects the number of samples removed after subtracting from the original 109 samples. For the θ(ψ) PTM, textural classification of core samples is shown in figure 7A (before excluding samples, $n = 108$; one sample had missing particle-size data and its texture could not be determined), and in figure 7B (after removing the maximum number of samples with errors and missing data, $n = 51$). For the K_{sat} PTM, textural classification of core samples is shown in figure 8A (before excluding samples with errors, $n = 96$; 13 samples had missing data) and in figure 8B (after removing the maximum number of samples with errors and missing data, $n = 78$).
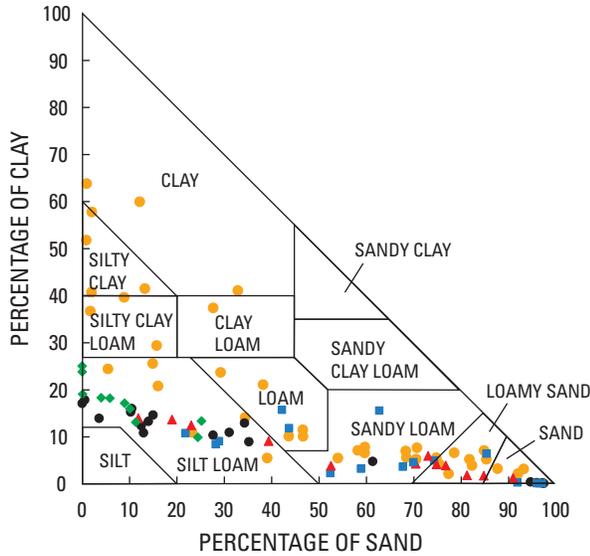
A. All potential core samples, except those with missing data (n = 96).
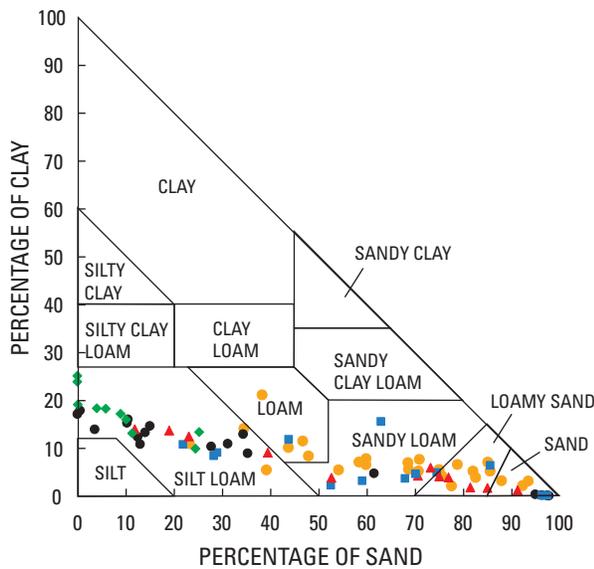
B. Core samples used in model development, shown with the maximum number removed due to errors and missing data (n = 78).

EXPLANATION

DATA SOURCE

- RWMC: Magnuson (2002, written commun.)
- RWMC: McElroy and Hubbell (1990)
- RWMC: Perkins and Nimmo (2000)
- VZRP: Perkins (2003)
- VZRP: Winfield (2003)

RWMC, Radioactive Waste Management Complex

VZRP, Vadose Zone Research Park

Figure 8. Textural classification of core samples (based on the system of the U.S. Department of Agriculture) used to develop a property-transfer model for estimating saturated hydraulic conductivity (K_{sat}).

Linearity Between Explanatory and Response Variables

Linearity between explanatory and response variables is an important assumption in applying the multiple linear-regression approach. Multiple linear regression is a correlation-based method, where measuring the strength of relations between variables assumes linearity. The strength of these relations will be underestimated if nonlinearity is present. The most direct way of correcting nonlinearity is by transforming one or both of the variables (Hair and others, 1998). Data transformations can improve the correlation between variables and correct violations of multivariate assumptions. Linearity was evaluated by examining scatter plots, using Pearson's *r*, and examining residual plots from preliminary regression analyses. Nonlinear trends identified in the scatter plots and residual plots were corrected by transforming the explanatory variable, the response variable, or both.

By examining scatter plots (not shown), and from the Pearson's *r* values (tables 6 and 7), several explanatory variables (M_g, σ_g, d_{50}, C_u, and percent gravel) appeared to vary nonlinearly with the response variables (θ_{sat}, ψ_o, λ, and K_{sat}). Based on the curvature of the relations between the explanatory and response variables, transformations such as the square root, inverse, or logarithm were selected as candidates for correcting the nonlinearity (Hair and others, 1998). Because the values of M_g, d_{50}, and C_u spanned approximately two orders of magnitude for the $\theta(\psi)$ models, and more than three orders of magnitude for the K_{sat} models, a logarithmic transformation was chosen to establish linearity. The spread in particle sizes is described by both σ_g and C_u. Although σ_g varied by less than an order of magnitude for the $\theta(\psi)$ and K_{sat} models, a logarithmic transformation also was applied to σ_g to provide a better comparison with $\log(C_u)$ in the comparison of competing regression models. Percent gravel varied by less than one order of magnitude. By definition, the textural-class percentages are constrained to vary between zero and 100 percent; the other particle-size parameters have no such lower and upper limits. Several core samples had zero gravel content, causing nonlinear patterns in the scatter plots. Transformation of percent gravel using the logarithm, inverse, or square root was not possible due to the zero values.

After transforming M_g, σ_g, d_{50}, and C_u, preliminary regression analyses were done and residuals for models A, B, and C were examined for each response variable. If residuals displayed heteroscedasticity, then the response variable was transformed. This procedure typically corrected all nonlinear effects in the scatter plots and in the residual plots. The response variables K_{sat}, ψ_o, and λ showed nonlinear relations with the explanatory variables. The logarithmic,

square-root, and inverse transformations were tried based on the shape of the relation between K_{sat}, ψ_o, and λ and the explanatory variables observed in the scatter plots (Hair and others, 1998). Because transformation of the response variable changes the model units, transformed models cannot be compared directly using the R^2_{adj}, *RMSE*, or other measures of model fit (Helsel and Hirsch, 1992). Instead the residuals were examined to determine the best transformation of K_{sat}, ψ_o, and λ to implement. K_{sat} varied over five orders of magnitude, ψ_o varied over three, and λ varied over one. The transformation providing the best residual behavior for K_{sat} and ψ_o was the logarithmic transformation. After the logarithmic transformation was applied, values of K_{sat} and ψ_o spanned less than an order of magnitude and the linearity of

K_{sat} and ψ_o with the explanatory variables greatly improved. For λ, the inverse and logarithmic transformations produced the best linear patterns in the scatter plots, but nonlinearity was not entirely corrected. The studentized residual patterns were similar for the logarithmic and inverse transformations, typically with slight heteroscedasticity existing for smaller predicted λ values. Applying the logarithmic transformation to λ improved the linearity with the explanatory variables, corrected most of the heteroscedasticity in the residual plots, and provided the most normally distributed residuals. An added benefit of using the logarithm of λ is that predicted values cannot be negative, which is a restriction on λ in the Rossi-Nimmo (1994) junction model.

Table 6. Correlation matrix between water-retention parameters and bulk-physical properties, without transformation of selected variables.

[The statistical relation between parameters listed in the header and first column is indicated by Pearson's correlation coefficient (r). Values of r can vary from -1 to +1, indicating parameters are either perfectly negatively or positively correlated. Values near zero indicate no correlation between parameters. Calculations were made using a subset of the original 109 core samples that excluded the samples with missing values and the Magnuson (2002) data, for a total of 57 samples. **Abbreviations:** g/cm³, gram per cubic centimeter; mm, millimeter; cm³/cm³, cubic centemeter per cubic centimeter; cm-water, centimeter of water]

	Bulk density, ρ_{bulk} (g/cm³)	Particle density, ρ_{part} (g/cm³)	Gravel (percent)	Sand (percent)	Silt (percent)	Clay (percent)	Geometric mean particle diameter, M_g (mm)	Geometric particle-size standard deviation, σ_g	Median particle diameter, d_{50} (mm)	Uniformity coefficient, C_u
Saturated water content, θ_{sat} (cm³/cm³)	-0.568	-0.046	-0.247	0.283	-0.250	-0.256	0.016	-0.036	0.074	0.047
Scaling parameter for matric pressure, ψ_o (-cm-water)	0.346	0.112	-0.097	-0.092	0.087	0.136	-0.091	-0.292	-0.134	-0.178
Curve-shape parameter, λ	-0.202	-0.057	0.221	0.490	-0.484	-0.505	0.721	-0.538	0.572	-0.364
Bulk density, ρ_{bulk} (g/cm³)	1	-0.058	0.063	-0.268	0.233	0.334	-0.320	-0.034	-0.399	-0.125
Particle density, ρ_{part} (g/cm³)		1	-0.039	-0.460	0.435	0.479	-0.221	-0.289	-0.256	-0.310
Gravel (percent)			1	0.201	-0.276	-0.271				
Sand (percent)				1	-0.992	-0.901				
Silt (percent)					1	0.858				
Clay (percent)						1				
Geometric mean particle diameter, M_g (mm)							1	-0.229		
Geometric particle-size standard deviation, σ_g								1		
Median particle diameter, d_{50} (mm)									1	0.217
Uniformity coefficient, C_u										1

Table 7.   Correlation matrix between saturated hydraulic conductivity and bulk-physical properties, without transformation of selected variables.

[The statistical relation between parameters listed in the header and first column is indicated by Pearson's correlation coefficient (r). Values of r can vary from -1 to +1, indicating parameters are either perfectly negatively or positively correlated. Values near zero indicate no correlation between parameters. Calculations were made using a subset of the original 109 core samples that excluded the samples with missing values, for a total of 96 samples. **Abbreviations:** g/cm^3, gram per cubic centimeter; mm, millimeter; cm/s, centimeter per second]

	Bulk density, ρ_{bulk} (g/cm^3)	Particle density, ρ_{part} (g/cm^3)	Gravel (percent)	Sand (percent)	Silt (percent)	Clay (percent)	Geometric mean particle diameter, M_g (mm)	Geometric particle-size standard deviation, σ_g	Median particle diameter, d_{50} (mm)	Uniformity coefficient, C_u
Saturated hydraulic conductivity, K_{sat} (cm/s)	-0.138	-0.032	0.326	0.132	-0.261	-0.148	0.250	0.127	0.210	0.077
Bulk density, ρ_{bulk} (g/cm^3)	1	-0.199	0.282	-0.311	0.013	0.379	0.216	0.177	0.231	0.039
Particle density, ρ_{part} (g/cm^3)		1	-0.160	-0.003	0.305	-0.408	-0.025	-0.328	-0.011	-0.155
Gravel (percent)			1	-0.045	-0.394	-0.213				
Sand (percent)				1	-0.819	-0.662				
Silt (percent)					1	0.381				
Clay (percent)						1				
Geometric mean particle diameter, M_g (mm)							1	0.178		
Geometric particle-size standard deviation, σ_g								1		
Median particle diameter, d_{50} (mm)									1	0.107
Uniformity coefficient, C_u										1

The effect of transforming the variables is reflected in Pearson's r values calculated before and after transformation by comparing tables 6 and 8 for the $\theta(\psi)$ models and tables 7 and 9 for the K_{sat} models. To simplify the discussion of correlation between variable pairs, the r values reported between each of the $\theta(\psi)$ parameters and the explanatory variables were based on 57 core samples (excluding the 51 Magnuson (written commun., 2002) samples and the sample with missing particle-size data from McElroy and Hubbell (1990)). The r values between K_{sat} and the explanatory variables were based on 96 core samples (excluding the 13 samples with missing particle-size or K_{sat} values). Similar r values were obtained using the final number of samples included in models A, B, and C after removing erroneous values. After applying a logarithmic transformation to log(ψ_o), the correlations with most textural-class parameters greatly improved. Prior to transformation, the r values between ψ_o and percent sand, percent silt, and percent clay were -0.092, 0.087, 0.136, respectively (table 6), reflecting the nonlinear relations observed in the scatter plots. The r values between log(ψ_o) and percent sand, percent silt, and percent clay improved after transformation of ψ_o (r = -0.645, 0.629, and 0.653, respectively, (table 8)). Similar improvements in r values were observed after applying the logarithmic transformation to other response variables (K_{sat} and λ) and some of the explanatory variables (M_g, σ_g, d_{50}, C_u).

Table 8.   Correlation matrix between water-retention parameters and bulk-physical properties, with transformation of selected variables.

[The statistical relation between parameters listed in the header and first column is indicated by Pearson's correlation coefficient (r). Values of r can vary from -1 to +1, indicating parameters are either perfectly negatively or positively correlated. Values near zero indicate no correlation between parameters. Calculations were made using a subset of the original 109 core samples that excluded the samples with missing values and the Magnuson (2002) data, for a total of 57 samples. **Abbreviations:** g/cm^3, gram per cubic centimeter; mm, millimeter; cm^3/cm^3, cubic centimeter per cubic centimeter; cm-water, centimeter of water]

	Bulk density, ρ_{bulk} (g/cm^3)	Particle density, ρ_{part} (g/cm^3)	Gravel (percent)	Sand (percent)	Silt (percent)	Clay (percent)	log(Geometric mean particle diameter, M_g (mm))	log(Geometric particle-size standard deviation, σ_g)	log(Median particle diameter, d_{50} (mm))	log(Uniformity coefficient, C_u)
Saturated water content, θ_{sat} (cm^3/cm^3)	-0.568	-0.046	-0.247	0.283	-0.250	-0.256	0.173	-0.038	0.175	-0.068
log(Scaling parameter for matric pressure, ψ_o (-cm-water))	0.565	0.315	-0.196	-0.645	0.629	0.653	-0.699	-0.183	-0.708	-0.095
log(Curve-shape parameter, λ)	-0.168	0.017	0.213	0.420	-0.406	-0.481	0.461	-0.684	0.408	-0.662
Bulk density, ρ_{bulk} (g/cm^3)	1	-0.058	0.063	-0.268	0.233	0.334	-0.317	-0.007	-0.323	0.044
Particle density, ρ_{part} (g/cm^3)		1	-0.039	-0.460	0.435	0.479	-0.437	-0.224	-0.450	-0.223
Gravel (percent)			1	0.201	-0.276	-0.271				
Sand (percent)				1	-0.992	-0.901				
Silt (percent)					1	0.858				
Clay (percent)						1				
log(Geometric mean particle diameter, M_g (mm))							1	-0.140		
log(Geometric particle-size standard deviation, σ_g)								1		
log(Median particle diameter, d_{50} (mm))									1	-0.060
log(Uniformity coefficient, C_u)										1

Poor correlations between some variable pairs indicated either that the explanatory variables had no relation with the response variable or that the linearity assumption was poorly met after applying transformations. All response variables were poorly correlated with percent gravel because most core samples had zero gravel content. As a result, the distribution for percent gravel was asymmetric, with large positive skewness, causing points on the scatter plots to be clustered near zero. ρ_{part} was poorly correlated with θ_{sat} (r = -0.046), log(λ) (r = 0.017), and log(K_{sat}) (r = -0.204), with only a slightly better correlation with log(ψ_o) (r = 0.315). By examining the scatter plots, it was determined that the relation between ρ_{part} and the response variables was weak. This was an early indication that ρ_{part} would not help explain the variation in θ_{sat}, log(λ), and log(K_{sat}) during the regression analyses. The implications for poor correlations between some variable pairs are that the coefficients may become unstable or may not be found significantly different from zero when in fact they are (incorrect significance testing). Regression models with poor correlations between variable pairs may be less desirable to use than models with explanatory variables better able to explain the variation in the response variable of interest. Assumptions of multiple linear regression are, on rare occasions, perfectly met. In this study, the explanatory variables for models A, B, and C were pre-selected to simplify comparison of different particle-size parameter sets; therefore, some variable pairs with low correlations were included in the regression analyses. Explanatory variables that are poorly correlated with the response variables will not likely contribute to explaining variation in response variables. Goodness-of-fit values likely would increase or decrease only slightly by including these explanatory variables in the regression variates.

Table 9.   Correlation matrix between saturated hydraulic conductivity and bulk-physical properties, with transformation of selected variables.

[The statistical relation between parameters listed in the header and first column is indicated by Pearson's correlation coefficient (r). Values of r can vary from -1 to +1, indicating parameters are either perfectly negatively or positively correlated. Values near zero indicate no correlation between parameters. Calculations were made using a subset of the original 109 core samples that excluded the samples with missing values, for a total of 96 samples. **Abbreviations:** g/cm^3, gram per cubic centimeter; mm, millimeter; cm/s, centimeter per second]

	Bulk density, ρ_{bulk} (g/cm^3)	Particle density, ρ_{part} (g/cm^3)	Gravel (percent)	Sand (percent)	Silt (percent)	Clay (percent)	log(Geometric mean particle diameter, M_g (mm))	log(Geometric particle-size standard deviation, σ_g)	log(Median particle diameter, d_{50} (mm))	log(Uniformity coefficient, C_u)
log(Saturated hydraulic conductivity, K_{sat} (cm/s))	-0.035	-0.204	0.369	0.655	-0.735	-0.476	0.728	0.270	0.710	0.197
Bulk density, ρ_{bulk} (g/cm^3)	1	-0.199	0.282	-0.311	0.013	0.379	-0.131	0.199	-0.151	0.165
Particle density, ρ_{part} (g/cm^3)		1	-0.160	-0.003	0.305	-0.408	-0.010	-0.347	0.063	-0.162
Gravel (percent)			1	-0.045	-0.394	-0.213				
Sand (percent)				1	-0.819	-0.662				
Silt (percent)					1	0.381				
Clay (percent)						1				
log(Geometric mean particle diameter, M_g (mm))							1	0.316		
log(Geometric particle-size standard deviation, σ_g)								1		
log(Median particle diameter, d_{50} (mm))									1	0.279
log(Uniformity coefficient, C_u)										1

Collinearity Between Explanatory Variables

Because collinearity affects the size of R^2, and can produce errors in coefficient magnitudes or signs, or instabilities in coefficient values, assessing collinear relations among explanatory variables is needed to understand the results of the multiple linear regression analyses. Collinearity was evaluated by using scatter plots, Pearson's r, and VIF values. Scatter plots between explanatory variables in model A indicated that percent sand was collinear with both percent silt and percent clay, and that percent silt was collinear with percent clay. This is true because any given textural-class percentage can be calculated by subtracting the sum of the remaining variables from 100 percent. For the $\theta(\psi)$ PTM, Pearson's r between percent sand and percent silt was -0.992; Pearson's r between percent sand and percent clay was -0.901; and Pearson's r between percent silt and percent clay was 0.858 (table 8). For similar reasons, the degree of correlation between percent sand and percent silt and between percent sand and percent clay was also high for the K_{sat} PTM (table 9). The correlation of percent gravel with the other textural-class percentages is smaller because a large number of core samples had zero gravel content. For the $\theta(\psi)$ PTM and the K_{sat} PTM, Pearson's r values indicated a poor degree of correlation among the explanatory variables in models B (ρ_{bulk}, M_g, and σ_g) and C (ρ_{bulk}, d_{50}, and C_u). VIF indicates the degree of multicollinearity in a model, which can come from the combined effect of two or more explanatory variables in the regression variate. Calculations of VIF for the $\theta(\psi)$ PTM were based on 57 core samples, excluding the 51 Magnuson (written commun., 2002) samples and the sample with missing particle-size data from McElroy and Hubbell (1990). For K_{sat}, 96 samples were used to calculate VIF values, after excluding the 13 samples with missing particle-size or K_{sat} values. For model A and the $\theta(\psi)$ PTM, VIF values ranged from 286 to 38,600 for percent gravel, sand, silt, and clay (table 10).

For the K_{sat} model, VIF values ranged from 124 to 605 for percent gravel, sand, silt, and clay. For VIF much greater than 10, results indicate a high degree of multicollinearity among the textural-class parameters. For models B and C, the VIF values ranged from 1.09 to 1.59 for explanatory variables in the $\theta(\psi)$ PTM and from 1.07 to 1.33 for those in the K_{sat} PTM (table 10), indicating a lack of multicollinearity among the explanatory variables. Because ρ_{bulk} does not relate directly to particle size, lack of multicollinearity is indicated by the low VIF values between ρ_{bulk} and the explanatory variables of models A, B, and C (ranging from 1.19 to 1.40 for the $\theta(\psi)$ PTM and from 1.10 to 1.40 for the K_{sat} PTM). The VIF values for ρ_{part} also were low because ρ_{part} has no relation with particle size.

Collinearity is acceptable as long as the prediction is made within the range of data values used in model development (Helsel and Hirsch, 1992). However, because the PTMs presented in this report will potentially be used to predict a wide range of sediment textures at the INEEL, beyond the range of textures in the calibration data, collinearity between textural-class percentages was deemed an undesirable aspect of model A. The sediment texture for the $\theta(\psi)$ models ranged from sands to silt loams (fig. 7B), which does not include materials with higher clay content found at the INEEL site. Solutions for correcting multicollinearity between variables in model A, but not used in this report, include centering the data, eliminating variables, collecting additional data, or performing a more sophisticated type of regression, such as ridge, Bayesian, or weighted least squares (Helsel and Hirsch, 1992; Hair and others, 1998).

Table 10. Variance inflation factors for explanatory variables used to develop property-transfer models for water retention and saturated hydraulic conductivity.

[VIF, variance inflation factor, measures multicollinearity between explanatory variables in a particular regression model. To calculate VIF, each explanatory variable is regressed against the remaining explanatory variables in the model, and the goodness-of-fit, indicated by the coefficient of determination (R^2), is calculated. VIF equals $1/(1-R^2)$. VIF less than 10 indicates little or no multicollinearity between the explanatory variable of interest and the rest of the variables used in the regression variate. VIF greater than 10 indicates the variable of interest is correlated with one or more of the other variables. Calculations were made using a subset of the original 109 core samples that excluded the samples with missing values, for a total of 57 samples for the water-retention models and a total of 96 samples for the saturated hydraulic conductivity models.]

Explanatory variable	Water-retention ($\theta(\psi)$) models		Saturated hydraulic conductivity (K_{sat}) models	
	Coefficient of determination (R^2)	Variance inflation factor (VIF)	Coefficient of determination (R^2)	Variance inflation factor (VIF)
Model A				
Bulk density, ρ_{bulk}	0.284	1.40	0.287	1.40
Particle density, ρ_{part}	.304	1.44	.436	1.77
Gravel	.997	286	.993	134
Sand	1.000	38,600	.998	605
Silt	1.000	27,400	.998	420
Clay	1.000	1,580	.992	124
Model B				
Bulk density, ρ_{bulk}	0.165	1.20	0.095	1.11
Particle density, ρ_{part}	.324	1.48	.145	1.17
log(Geometric mean particle diameter, M_g)	.372	1.59	.146	1.17
log(Geometric particle-size standard deviation, σ_g)	.137	1.16	.250	1.33
Model C				
Bulk density, ρ_{bulk}	0.157	1.19	0.092	1.10
Particle density, ρ_{part}	.308	1.45	.063	1.07
log(Median particle, diameter, d_{50})	.348	1.53	.124	1.14
log(Uniformity coefficient, C_u)	.084	1.09	.142	1.17

Water-Retention Property-Transfer Model

Because θ varies nonlinearly with ψ, the Rossi-Nimmo (1994) junction model was used to characterize the $\theta(\psi)$ curve. The Rossi-Nimmo (1994) junction model requires three parameters (θ_{sat}, ψ_o, and λ) to describe the entire range of ψ; therefore, three separate multiple linear-regression equations are needed.

Saturated Water Content

Regression results for θ_{sat} were analyzed for overall goodness-of-fit, normality of residuals, homoscedasticity of residuals, and significance of coefficients. A summary of regression results for θ_{sat}, including values of R^2_{adj}, $RMSE$, PPCC, and residual-distribution skewness, for models A, B, and C are presented in table 11. The regression analyses were completed after removing as many as 57 core samples

identified as erroneous from the original total of 109 (see table 5). For samples included in models A, B, and C, θ_{sat} ranged from 0.3473 to 0.6151. Goodness-of-fit values (R^2_{adj}) were comparable for each model, ranging from 0.491 to 0.568, with model A having the highest value. The residuals were normally distributed for all models, as indicated by probability plots, PPCC values, and skewness values. PPCC values ranged from 0.989 to 0.992 compared to the critical value of 0.981. Residual distributions showed slight negative skewness, with values ranging from -0.05 to -0.20. Residual plots did not show heteroscedastic or nonlinear patterns (see example in fig. 9A for model C). Predicted θ_{sat} values are plotted against observed values in figure 10A for model C. The θ_{sat} values for the samples from Perkins and Nimmo (2000) were consistently overestimated, falling above the 1:1 line. Most similarly textured samples had higher θ_{sat} values than the Perkins and Nimmo (2000) samples, which may be a cause for θ_{sat} overestimation.

Table 11. Multiple linear-regression results comparing alternative representations of particle-size distribution for water-retention parameters and saturated hydraulic conductivity.

[Hydraulic parameters used as response variables in regression models included saturated water content, the scaling parameter for matric pressure, the water-retention curve-shape parameter, and saturated hydraulic conductivity. Three models (A, B, and C) with alternative forms of the particle-size distribution as input were compared for each hydraulic parameter. R^2_{adj} indicates the goodness-of-fit for the model, accounting for the number of explanatory variables and the number of observations; values close to 1 indicate that the variables successfully explain the variation in the hydraulic parameter. $RMSE$, in units of the hydraulic parameter, measures the variation in the predicted values of each hydraulic parameter, similar to the standard deviation of a variable around its mean. PPCC indicates the normality of a distribution by comparing it to the normal distribution (a straight line) on a probability plot. If PPCC is greater than the critical PPCC value, then the data are considered normally distributed. **Abbreviations:** R^2_{adj}, adjusted coefficient of determination; $RMSE$, root-mean-square-error; PPCC, probability plot correlation coefficient; ρ_{bulk}, bulk density; M_g, geometric mean particle diameter; σ_g, geometric particle-size standard deviation; d_{50}, median particle diameter; C_u, uniformity coefficient; cm^3/cm^3, cubic centimeter per cubic centimeter; cm-water, centimeter of water; cm/s, centimeter per second]

Model	Explanatory variables	R^2_{adj}	$RMSE$	PPCC for residual distribution	Critical PPCC	Skewness of residual distribution	Number of observations, n
Saturated water content, θ_{sat} (cm^3/cm^3)							
A	ρ_{bulk}, gravel, sand, silt, clay	0.568	0.043	0.989	0.981	-0.20	51
B	ρ_{bulk}, $\log(M_g)$, $\log(\sigma_g)$	.491	.047	.992	.981	-.06	52
C	ρ_{bulk}, $\log(d_{50})$, $\log(C_u)$	.491	.047	.992	.981	-.05	52
Scaling parameter for matric pressure, ψ_o (-cm-water)							
A	ρ_{bulk}, gravel, sand, silt, clay	0.667	0.406	0.993	0.982	-0.04	54
B	ρ_{bulk}, $\log(M_g)$, $\log(\sigma_g)$	.727	.367	.986	.982	-.33	55
C	ρ_{bulk}, $\log(d_{50})$, $\log(C_u)$	.693	.389	.987	.982	-.32	55
Curve-shape parameter, λ							
A	ρ_{bulk}, gravel, sand, silt, clay	0.192	0.230	0.990	0.982	0.23	54
B	ρ_{bulk}, $\log(M_g)$, $\log(\sigma_g)$	.667	.157	.993	.982	-.16	55
C	ρ_{bulk}, $\log(d_{50})$, $\log(C_u)$	.607	.170	.989	.982	-.05	55
Saturated hydraulic conductivity, K_{sat} (cm/s)							
A	ρ_{bulk}, gravel, sand, silt, clay	0.632	0.918	0.994	0.988	-0.19	91
B	ρ_{bulk}, $\log(M_g)$, $\log(\sigma_g)$	.606	.982	.991	.987	-.34	78
C	ρ_{bulk}, $\log(d_{50})$, $\log(C_u)$	.589	1.003	.989	.987	-.36	78
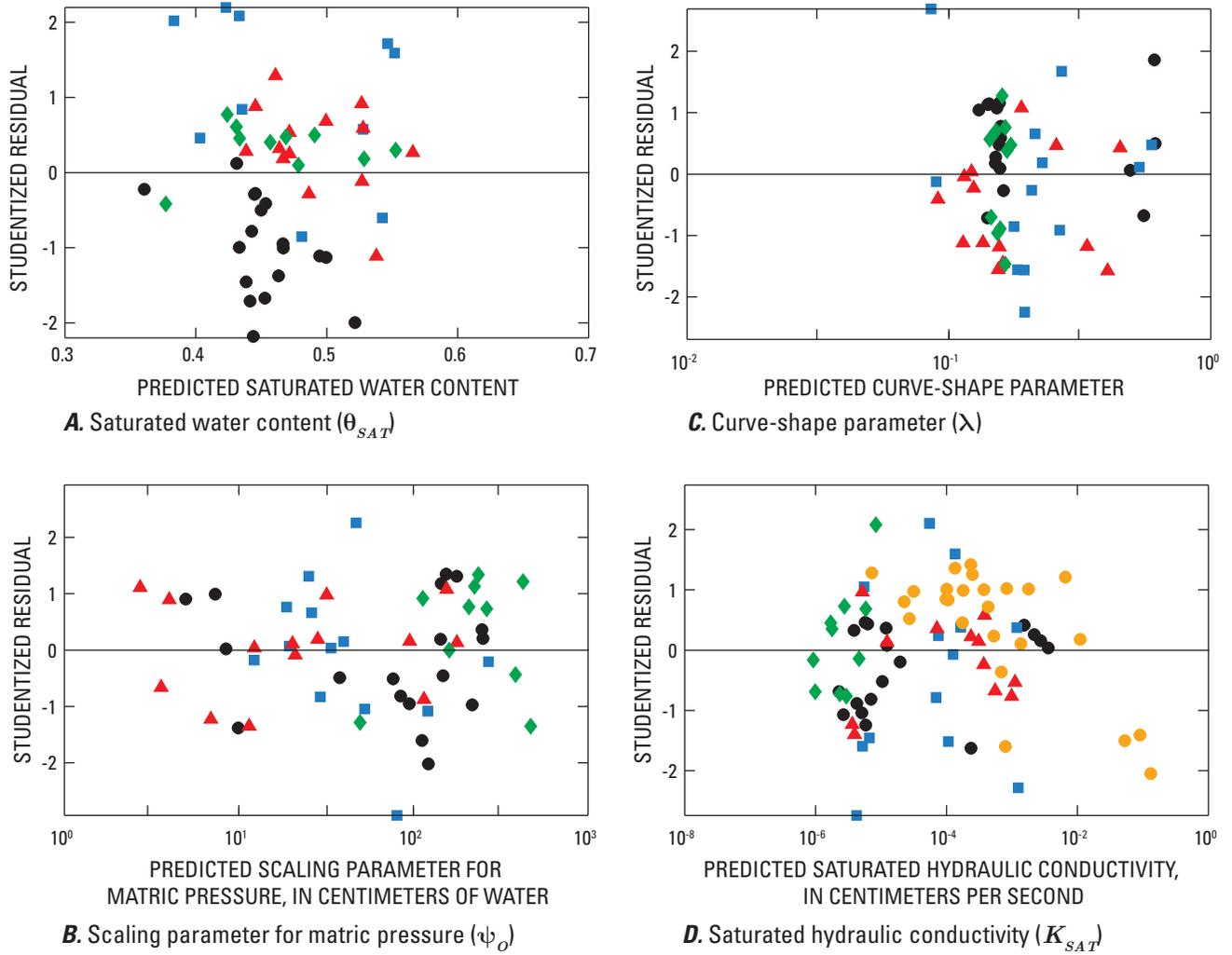
A. Saturated water content (θ_{SAT})

C. Curve-shape parameter (λ)

B. Scaling parameter for matric pressure (ψ_O)

D. Saturated hydraulic conductivity (K_{SAT})

EXPLANATION

DATA SOURCE

- ● RWMC: Magnuson (2002, written commun.)   ▲ VZRP: Perkins (2003)
- ■ RWMC: McElroy and Hubbell (1990)   ◆ VZRP: Winfield (2003)
- ● RWMC: Perkins and Nimmo (2000)   RWMC, Radioactive Waste Management Complex
- VZRP, Vadose Zone Research Park

Figure 9.   Studentized residuals and predicted hydraulic property values for multiple-linear regression model C (where the median particle diameter, uniformity coefficient, and bulk density were used as the explanatory variables).

Figure 10. Predicted and observed hydraulic property values for multiple-linear regression model C (where the median particle diameter, uniformity coefficient, and bulk density were used as the explanatory variables).

Because the residual behavior was equivalent among models A, B, and C, the best model for θ_{sat} was chosen as model A, which had the highest R^2_{adj} and lowest *RMSE*. However, model A was able to explain only about 8 percent more of the variation in θ_{sat}, as indicated by the values of R^2_{adj}, than the other two models. The null hypothesis test states that coefficients are significantly different from zero if their t-value magnitudes exceed the critical t-value for the specified α-level and number of observations in the model. The partial t-value magnitudes for M_g, σ_g, d_{50}, and C_u were less than the critical t-value of 2.011 (table 12). These variables did not help explain the variation in θ_{sat}, which is why the goodness-of-fit values for models B and C were nearly identical. For models A, B, and C, the partial t-value magnitudes for ρ_{bulk} ranged from -6.975 to -6.960, greatly exceeding the critical t-value of about 2.010 at $\alpha = 0.05$. ρ_{bulk} was the most significant variable for explaining the variation in θ_{sat}. This finding also was supported by the values of Pearson's *r* calculated between each explanatory variable and θ_{sat}, where ρ_{bulk} was the only explanatory variable that had at least a moderate correlation with θ_{sat} (r = -0.568; table 8). Textural-class percentages in model A are affected by multicollinearity, and are not highly

correlated with θ_{sat} (table 8). The geometric and graphical particle-size variables in models B and C do not correlate highly with θ_{sat}, which is a likely reason why their coefficients were not significantly different from zero.

Because θ_{sat} is related to total porosity, which can be calculated from both ρ_{bulk} and ρ_{part}, the effect of including ρ_{part} in models A, B, and C was evaluated. The R^2_{adj} value decreased by 0.4 percent and 0.5 percent for models B and C, respectively, and did not change for model A. From the scatter plots and Pearson's *r* (table 8), ρ_{part} was not strongly correlated with θ_{sat}. Additionally, ρ_{part} was poorly correlated with ψ_o, λ, and K_{sat} (tables 7 and 9). Therefore, ρ_{part} was excluded from the regression analyses of all response variables.

Scaling Parameter for Matric Pressure

The regression results for ψ_o are listed in tables 11 and 13. The regression analyses were based on a subset of the original 109 core samples, with the final number of samples ranging from 54 to 55. Values of ψ_o ranged from 2.12 cm-water for a sandy loam to 1,207.78 cm-water for a silt loam.

Table 12.   Multiple linear-regression model coefficients, confidence limits, standard errors, and partial t-values for the water-retention parameter, saturated water content.

[Three regression models (A, B, and C) with alternative forms of the particle-size distribution as input were compared. Regression coefficients and the upper and lower 95-percent confidence limits are listed for each explanatory variable in a given model. Standard error denotes the expected variation of a coefficient across multiple samples of the data from a given population. Smaller standard error values imply more reliable predictions (smaller confidence intervals for the coefficient). Partial t-value, used to test whether a coefficient is significantly different from zero, is calculated by dividing the coefficient by its standard error. If the magnitude of the partial t-value is greater than the critical t-value (calculated from the Student's t-distribution at a significance level of 0.05 and for the number of observations in the model), then the coefficient is significantly different from zero]

Explanatory variable	Coefficient	Lower boundary	Upper boundary	Standard error	Partial t-value	Critical t-value
			Model A			
Y-intercept	-5.0123	-12.6245	2.6001	3.780	-1.326	2.014
Bulk density, ρ_{bulk}	-.3974	-.5121	-.2826	.057	-6.973	
Gravel	.0604	-.0161	.1369	.038	1.590	
Sand	.0595	-.0170	.1359	.038	1.567	
Silt	.0641	-.0129	.1411	.038	1.677	
Clay	.1353	-.0502	.3209	.092	1.469	
			Model B			
Y-intercept	1.0029	0.8453	1.1605	0.078	12.796	2.011
Bulk density, ρ_{bulk}	-.4002	-.5156	-.2848	.057	-6.975	
log(Geometric mean particle diameter, M_g)	-.0130	-.0379	.0119	.012	-1.049	
log(Geometric particle-size standard deviation, σ_g)	-.0022	-.0872	.0827	.042	-.053	
			Model C			
Y-intercept	1.0063	0.8531	1.1595	0.076	13.207	2.011
Bulk density, ρ_{bulk}	-.3998	-.5153	-.2843	.057	-6.960	
log(Median particle diameter, d_{50})	-.0123	-.0362	.0115	.012	-1.039	
log(Uniformity coefficient, C_u)	-.0029	-.0372	.0314	.017	-.171	

The R^2_{adj} values were high for all models, ranging between 0.667 and 0.727. Model B had the highest R^2_{adj} (0.727) and smallest *RMSE* (0.367). The residuals for model A were the most normally distributed. Although the PPCC for model B was greater than the critical value (0.986 > 0.982), the residual distribution for model B was more negatively skewed than that of model A or C (with skewness of -0.33 for model B compared to -0.04 for model A and -0.32 of model C). For models B and C, the studentized residuals were homoscedastic (fig. 9*B* for model C). Observations from different data sets were not preferentially underestimated or overestimated, as shown by the example for model C in figure 10*B*.

The coefficients for each model were examined for significance from zero using the partial t-test. If the partial t-value magnitude exceeded the critical t-value calculated from Student's t-distribution at $\alpha = 0.05$, then the coefficient was significantly different from zero. As with θ_{sat}, only ρ_{bulk} was significant in explaining the variation in $\log(\psi_o)$ for model A (table 13). The correlation between ρ_{bulk} and $\log(\psi_o)$ was moderate, with $r = 0.565$ (table 8). However, compared with the θ_{sat} results for model A, the textural parameters—percent sand, silt, and clay—are at least moderately correlated with $\log(\psi_o)$, with values of Pearson's r of -0.645, 0.629, and 0.653, respectively (table 8). The coefficients for these parameters did not pass the partial t-test, indicating the influence of multicollinearity on model A. The coefficients for ρ_{bulk}, M_g, and σ_g were significantly different than zero because the partial t-value magnitudes were greater than the critical t-value of 2.008. $\log(\psi_o)$ was moderately correlated with $\log(M_g)$ ($r = -0.699$) and poorly correlated with $\log(\sigma_g)$ ($r = -0.183$). For model C, the coefficients for the intercept term, ρ_{bulk}, and d_{50} were significantly different from zero. ρ_{bulk} and d_{50} also showed the highest univariate correlations with $\log(\psi_o)$ for model C. Overall, model B had the greatest number of explanatory variables that would be helpful in predicting $\log(\psi_o)$.

Based on residual behavior and goodness-of-fit values, model B was chosen as the best model for predicting $\log(\psi_o)$. The residuals of model B were the most normally distributed according to the PPCC and did not display unequal variances when plotted against predicted values of $\log(\psi_o)$. Model B had the highest goodness-of-fit values, explaining 6 percent more of the variation in $\log(\psi_o)$ than model A and 3 percent more

Table 13. Multiple linear-regression model coefficients, confidence limits, standard errors, and partial t-values for the water-retention scaling parameter for matric pressure.

[Three regression models (A, B, and C) with alternative forms of the particle-size distribution as input were compared. Regression coefficients and the upper and lower 95-percent confidence limits are listed for each explanatory variable in a given model. Standard error denotes the expected variation of a coefficient across multiple samples of the data from a given population. Smaller standard error values imply more reliable predictions (smaller confidence intervals for the coefficient). Partial t-value, used to test whether a coefficient is significantly different from zero, is calculated by dividing the coefficient by its standard error. If the magnitude of the partial t-value is greater than the critical t-value (calculated from the Student's t-distribution at a significance level of 0.05 and for the number of observations in the model), then the coefficient is significantly different from zero]

Explanatory variable	Coefficient	Lower boundary	Upper boundary	Standard error	Partial t-value	Critical t-value
			Model A			
Y-intercept	51.4054	-17.4564	120.2670	34.249	1.501	2.011
Bulk density, ρ_{bulk}	1.8613	.8042	2.9183	.526	3.540	
Gravel	-.5303	-1.2235	.1628	.345	-1.538	
Sand	-.5199	-1.2124	.1727	.344	-1.509	
Silt	-.4977	-1.1964	.2011	.348	-1.432	
Clay	-1.2932	-2.9783	.3920	.838	-1.543	
			Model B			
Y-intercept	-1.1183	-2.3143	0.0777	0.596	-1.877	2.008
Bulk density, ρ_{bulk}	1.4866	.6178	2.3554	.433	3.435	
log(Geometric mean particle diameter, M_g)	-.9035	-1.0956	-.7115	.096	-9.445	
log(Geometric particle-size standard deviation, σ_g)	-.8969	-1.5605	-.2332	.331	-2.713	
			Model C			
Y-intercept	-1.4080	-2.6350	-0.1810	0.611	-2.304	2.008
Bulk density, ρ_{bulk}	1.5344	.6139	2.4549	.459	3.346	
log(Median particle diameter, d_{50})	-.8394	-1.0330	-.6458	.096	-8.703	
log(Uniformity coefficient, C_u)	-.1510	-.4347	.1328	.141	-1.068	

than model C. All coefficients for model B were significantly different from zero, indicating that all explanatory variables were important for describing the variation in $\log(\psi_o)$.

Curve-Shape Parameter

Multiple linear-regression analyses for $\log(\lambda)$ are summarized in tables 11 and 14 for models A, B, and C. After removing samples with erroneous or missing values (table 5), the final number of samples used in model A was 54, and the number used in models B and C was 55. Model B had the highest R^2_{adj} and smallest *RMSE*. The residuals also were the most normally distributed for this model, with a PPCC value of 0.993 compared to the critical PPCC value of 0.982, and a skewness value of -0.16. The residual plot for model B showed the least heteroscedasticity. Residual plots for models A and C showed some clustering of points at smaller values of predicted $\log(\lambda)$ (fig. 9C), and the residual plot for model A showed slight nonlinearity. Most of the λ values for the Perkins and Nimmo (2000) data set were underestimated (fig. 10C for model C). The coefficients for $\log(M_g)$, $\log(\sigma_g)$,

$\log(d_{50})$, and $\log(C_u)$ were significantly different from zero for models B and C; whereas for model A, only the coefficient for percent silt was significantly different from zero. From Pearson's r values (table 8), $\log(\sigma_g)$ and $\log(C_u)$ were most highly correlated with $\log(\lambda)$. Based on the comparison of regression results for models A, B, and C, model B was selected as the best model for predicting $\log(\lambda)$.

Saturated Hydraulic Conductivity Property-Transfer Model

The K_{sat} PTM was developed using the same core samples as the $\theta(\psi)$ PTM in addition to the 51 samples from Magnuson (written commun., 2002), for 96 potential samples (excluding the samples with missing K_{sat} and particle-size measurements). Twelve samples had missing K_{sat} values (10 samples from Magnuson (written commun., 2002) and two samples from Perkins (2003)) and one sample had partial particle-size data (USGS 118 at 173.48 to 174.39 m from McElroy and Hubbell (1990)); these samples were treated

Table 14.   Multiple linear-regression model coefficients, confidence limits, standard errors, and partial t-values for the water-retention curve-shape parameter.

[Three regression models (A, B, and C) with alternative forms of the particle-size distribution as input were compared. Regression coefficients and the upper and lower 95-percent confidence limits are listed for each explanatory variable in a given model. Standard error denotes the expected variation of a coefficient across multiple samples of the data from a given population. Smaller standard error values imply more reliable predictions (smaller confidence intervals for the coefficient). Partial t-value, used to test whether a coefficient is significantly different from zero, is calculated by dividing the coefficient by its standard error. If the magnitude of the partial t-value is greater than the critical t-value (calculated from the Student's t-distribution at a significance level of 0.05 and for the number of observations in the model), then the coefficient is significantly different from zero]

Explanatory variable	Coefficient	Lower boundary	Upper boundary	Standard error	Partial t-value	Critical t-value
Model A						
Y-intercept	38.0186	-0.9624	76.9996	19.387	1.961	2.011
Bulk density, ρ_{bulk}	.1387	-.4596	.7371	.298	.466	
Gravel	-.3881	-.7805	.0043	.195	-1.989	
Sand	-.3867	-.7788	.0053	.195	-1.983	
Silt	-.4099	-.8055	-.0144	.197	-2.084	
Clay	-.9143	-1.8683	.0396	.474	-1.927	
Model B						
Y-intercept	0.2489	-0.2615	0.7594	0.254	0.979	2.008
Bulk density, ρ_{bulk}	.0799	-.2909	.4507	.185	.433	
log(Geometric mean particle diameter, M_g)	.2029	.1209	.2848	.041	4.970	
log(Geometric particle-size standard deviation, σ_g)	-1.2227	-1.5059	-.9394	.141	-8.666	
Model C						
Y-intercept	-0.0411	-0.5777	0.4955	0.267	-0.154	2.008
Bulk density, ρ_{bulk}	.0974	-.3051	.4999	.201	.486	
log(Median particle diameter, d_{50})	.1925	.1079	.2772	.042	4.565	
log(Uniformity coefficient, C_u)	-.4910	-.6151	-.3669	.062	-7.944	

as missing data and subtracted from the original set of 109 samples. K_{sat} values of the 96 samples ranged from 1.13×10^{-8} to 8.41×10^{-2} cm/s (table 2), with textures ranging from clays to sands (fig. 8A). Thirty six of the 96 samples were classified as silt loams, with the second greatest number of samples (19) classified as sandy loams. As discussed above, five additional core samples were removed from model A due to either inadequacies in measured particle-size distributions or errors in measured K_{sat} and ρ_{part} values. An additional 18 samples were removed from models B and C for similar reasons. The total number of core samples used to develop model A was 91, with 78 samples used to develop models B and C. The range of textures for models B and C is shown on the textural triangle in figure 8B.

Regression results for each K_{sat} model are summarized in tables 11 and 15. Model A had the highest R^2_{adj} value (0.632), but no coefficients were significantly different from zero due to the multicollinearity between the textural-class parameters. The residuals for Model A were the most normally distributed, with a PPCC value of 0.994 compared to the critical value of 0.988, and a skewness of -0.19. Model B had the second highest R^2_{adj} value (0.606) and ranked second best in terms of

residual normality (PPCC of 0.991 compared to the critical value of 0.987, and skewness of -0.34). Coefficients for the intercept term and $\log(M_g)$ were significantly different from zero. The correlation between $\log(K_{sat})$ and $\log(M_g)$ ($r = 0.728$; table 9) was highest among all the explanatory variables in model B. Model C had the lowest R^2_{adj} value (0.589). Although the normality of the residuals was found to be good, with a PPCC of 0.989 compared to the critical PPCC value of 0.987, the residual distribution was slightly bimodal. Only the intercept term and $\log(d_{50})$ had coefficients that passed the partial t-test for significance from zero. $\log(d_{50})$ also was more highly correlated with $\log(K_{sat})$ ($r = 0.710$; table 9) than ρ_{bulk} or $\log(C_u)$. Studentized residual variances, plotted against the predicted K_{sat} values, were nearly random and equal for all models (fig. 9D for model C), with the exception of a few extreme values exceeding the t-value threshold of about 2.0. Trends observed in residual plots were basically linear. Predicted values of K_{sat} are shown plotted against observed values for model C in figure 10D. The K_{sat} values for the Magnuson samples tended to be underestimated, likely due to the large number of samples with high clay contents and high

Table 15. Multiple linear-regression model coefficients, confidence limits, standard errors, and partial t-values for saturated hydraulic conductivity.

[Three regression models (A, B, and C) with alternative forms of the particle-size distribution as input were compared. Regression coefficients and the upper and lower 95-percent confidence limits are listed for each explanatory variable in a given model. Standard error denotes the expected variation of a coefficient across multiple samples of the data from a given population. Smaller standard error values imply more reliable predictions (smaller confidence intervals for the coefficient). Partial t-value, used to test whether a coefficient is significantly different from zero, is calculated by dividing the coefficient by its standard error. If the magnitude of the partial t-value is greater than the critical t-value (calculated from the Student's t-distribution at a significance level of 0.05 and for the number of observations in the model), then the coefficient is significantly different from zero]

Explanatory variable	Coefficient	Lower boundary	Upper boundary	Standard error	Partial t-value	Critical t-value
Model A						
Y-intercept	-14.2998	-29.4376	0.8379	7.614	-1.878	1.988
Bulk density, ρ_{bulk}	.6625	-.5308	1.8557	.600	1.104	
Gravel	.1123	-.0381	.2627	.076	1.484	
Sand	.1127	-.0387	.2641	.076	1.480	
Silt	.0712	-.0788	.2212	.075	.944	
Clay	.0826	-.0685	.2336	.076	1.087	
Model B						
Y-intercept	-1.5121	-3.4163	0.3920	0.956	-1.582	1.993
Bulk density, ρ_{bulk}	-.0500	-1.2985	1.1985	.627	-.080	
log(Geometric mean particle diameter, M_g)	1.8423	1.4862	2.1984	.179	10.308	
log(Geometric particle-size standard deviation, σ_g)	-.2906	-1.3644	.7833	.539	-.539	
Model C						
Y-intercept	-1.7690	-3.6640	0.1260	0.951	-1.860	1.993
Bulk density, ρ_{bulk}	.0794	-1.1796	1.3384	.632	.126	
log(Median particle diameter, d_{50})	1.7507	1.4155	2.0858	.168	10.408	
log(Uniformity coefficient, C_u)	-.3274	-.8314	.1765	.253	-1.295	

K_{sat} values compared with the rest of the data. Based on the high value of R^2_{adj}, and the normality and homoscedasticity of the residuals, model A was identified as the best model for predicting K_{sat}.

Model Discrimination

Several tests were done to keep the use of the $\theta(\psi)$ and K_{sat} models practical by limiting explanatory variables to one common set of particle-size parameters and determining the sensitivity of the estimated $\theta(\psi)$ and K_{sat} values to the way the particle-size distribution was represented. For the θ_{sat}, ψ_o, and λ models, R^2_{adj} values were not consistently higher for one common set of particle-size parameters. The set of three regression models for θ_{sat}, ψ_o, and λ that produced the highest R^2_{adj} values was model combination A-B-B (textural-class percentages (A) for θ_{sat}, and geometric statistics (B) for ψ_o and λ). These models also showed the best behavior in terms of meeting the assumptions of multiple linear regression. Model A had the highest R^2_{adj} and displayed the best model behavior for K_{sat}. Because the combined ability of the θ_{sat}, ψ_o, and λ models in estimating the $\theta(\psi)$ curve is of more interest than obtaining the highest goodness-of-fit values for individual parameters, four model combinations were tested based on: (1) highest R^2_{adj} values (A-B-B), (2) textural-class percentages (A-A-A), (3) geometric statistics (B-B-B), and (4) graphical statistics (C-C-C). Model discrimination then was completed by (1) calculating the *RMSE* between the observed and predicted $\theta(\psi)$ curve for each core sample, (2) comparing the *RMSE* distributions of the four competing model combinations, and (3) evaluating the best model combination based on ease of model use, best model behavior, highest values of R^2_{adj}, and best representation of the particle-size distribution to use for development of a theoretical PTM.

For each core sample, predicted values of θ_{sat}, ψ_o, and λ were calculated using the appropriate regression equations developed for models A, B, and C. For each of the four model combinations tested, the predicted values of θ_{sat}, ψ_o, and λ were used as input to the Rossi-Nimmo (1994) junction model (eqs. 1 to 3) to produce predicted $\theta(\psi)$ curves for each core sample. The $\theta(\psi)$ curve could not be estimated if any value of θ_{sat}, ψ_o, or λ was missing for a particular core sample. For this reason, the number of estimated curves per model combination was constrained to the minimum number of core samples used to develop the θ_{sat}, ψ_o, and λ models as a whole. For model combination A-B-B, 50 samples were used, and for model combination A-A-A, 51 samples were used. For model combinations B-B-B and C-C-C, 52 samples were used in the comparisons. Because ψ_o and λ were log-transformed, and the range of measured θ_{sat} values was not close to zero, predicted values of the hydraulic parameters were always positive. Negative values would cause errors in the $\theta(\psi)$ curve calculated with the Rossi-Nimmo (1994) junction model.

Due to the difficulty in visually assessing the quality of predicted $\theta(\psi)$ curves for each model combination (which would require up to 52 comparisons per combination), R^2 and *RMSE* values were used to assess goodness-of-fit between observed and predicted $\theta(\psi)$ curves. For each core sample, the observed $\theta(\psi)$ curve was defined using the measured θ_{sat} value and the two parameters (ψ_o and λ) obtained by fitting the Rossi-Nimmo (1994) junction model (eqs. 1 to 3) to the $\theta(\psi)$ measurements. The R^2 and *RMSE* values originally were computed from θ values at the measured ψ values, however, goodness-of-fit values were found to be sensitive to the range and number of measured $\theta(\psi)$ points. Among core samples, the measured $\theta(\psi)$ points spanned different ranges of ψ and varied in number. R^2, calculated between the observed and predicted θ values at the measured ψ values, was highly sensitive to the range and number of $\theta(\psi)$ points. For some core samples, the R^2 value was high (0.90 or greater) even when the observed and estimated curves visually appeared quite different. To place samples on an equal basis for comparison, observed and predicted θ values were calculated at 1,000 logarithmically-spaced ψ points to define an approximately smooth curve, and R^2 and *RMSE* values were recalculated. R^2 values increased using the finer resolution of ψ points and still falsely indicated goodness-of-fit for some samples. Although *RMSE* differs from R^2 because it is expressed in units of the dependent variable (θ) and does not have a fixed scale like R^2, which varies from 0 to 1, *RMSE* was chosen as the best goodness-of-fit indicator for comparing the four model combinations because R^2 did not always indicate true goodness-of-fit.

Predicted $\theta(\psi)$ curves for each model combination are shown in figure 11 for two core samples of silt loam texture with contrasting goodness-of-fit values. For model combination C-C-C, the *RMSE* between observed and predicted curves for core sample ICPP-SCI-V-214 at 56.24 to 56.33 m from Winfield (2003) (fig. 11*A*) was 0.012, indicating a good fit, while the *RMSE* for core sample UZ98-2 at 49.89 to 49.99 m from Perkins and Nimmo (2000) (fig. 11*B*) was 0.043, indicating a poor fit. The overestimation of θ_{sat} inflated the *RMSE* value for UZ98-2 at 49.89 to 49.99 m, indicating the importance of accurately estimating θ_{sat} because it controls the relative position of the $\theta(\psi)$ curve along the θ axis.

RMSE distributions for each model combination were examined to determine the best particle-size parameter set to implement based on the distribution with the smallest *RMSE* values. In figure 12, the *RMSE* distributions for each model combination are presented as a boxplot. The box represents the interquartile range for each R^2 distribution, where the upper box limit is the 75th percentile and the lower limit is the 25th percentile. The median of the distribution (50th percentile) is drawn as a line across the box. Skewness is indicated by the relative position of the median line within the box, with a perfect normal distribution indicated by a line centered in

the box. The "whiskers" extend from the box edge to the data points that lie within 1.5 times the interquartile range. Any points that lie beyond the ends of the whiskers, plotted as circles, are outlying points. If there are no outlying points, a dot is placed at the end of the left whisker. The range of *RMSE* values was similar between each model combination, varying from 0.001 to 0.064 (fig. 12). The medians ranged from 0.015

to 0.021. These results indicated that the estimated $\theta(\psi)$ curves were insensitive to the way the particle-size distribution was represented. The distributions for model combinations A-B-B and B-B-B were nearly identical, indicating that the estimated θ_{sat} values for each core sample were insensitive to the regression equation used. This result supports the earlier observation that the particle-size parameters explain little

A. Core sample ICPP-SCI-V-214 at 56.24 to 56.33 meters (Perkins, 2003)

B. Core sample UZ98-2 at 49.89 to 49.99 meters (Perkins and Nimmo, 2000)
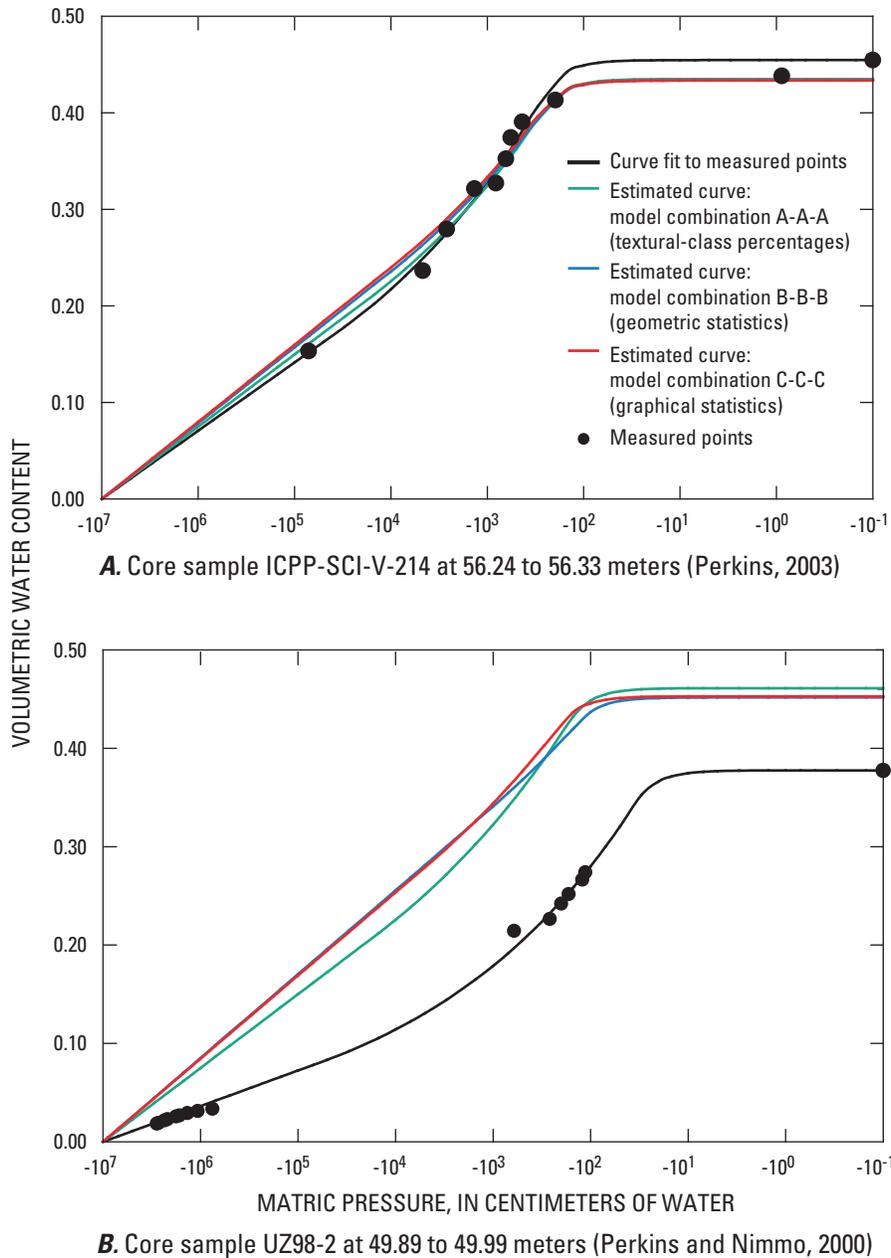
Figure 11.   Estimated water-retention ($\theta(\psi)$) curves for two silt loam core samples based on three multiple linear-regression models with different particle-size parameter sets as input.
The $\theta(\psi)$ curve was represented by the Rossi-Nimmo (1994) junction model.
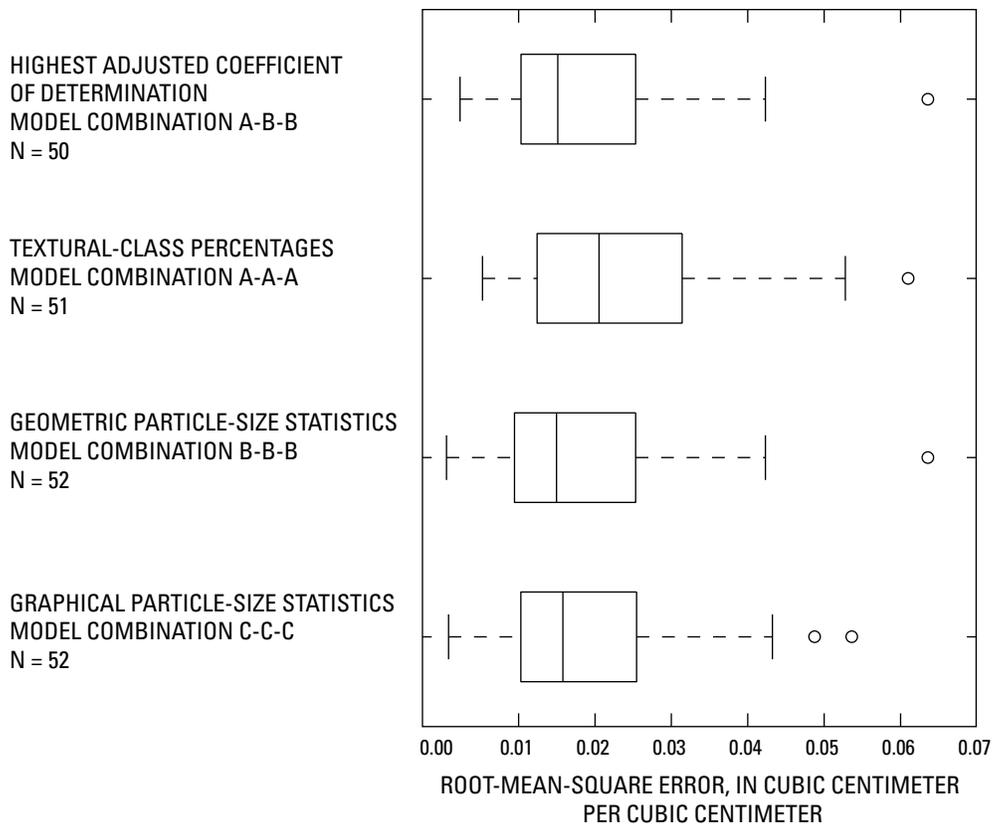
Figure 12.   Comparison of root-mean-square-error (*RMSE*) distributions for four combinations of multiple linear-regression models with different particle-size parameter sets as input.
RMSE was calculated for individual core samples between water-retention curves estimated from regression models and fitted to measured points. The water-retention curve was represented by the Rossi-Nimmo (1994) junction model.

of the variation in θ_{sat}, and that ρ_{bulk} is the most important explanatory variable for θ_{sat}. Model combination A-A-A possessed the largest interquartile range, or spread of *RMSE* values, likely because the λ model had such a low value of R^2_{adj} (0.192).

The *RMSE* distribution for model combination C-C-C was divided into groups to examine trends based on texture, source data, and location. The 52 core samples for this model combination were divided into groups based on the five U.S. Department of Agriculture textures (sand, loamy sand, sandy loam, loam, and silt loam) represented by the data (fig. 13*A*). The primary texture represented by the data was silt loam ($n = 29$), followed by sandy loam ($n = 10$). Sandy loam had the smallest median *RMSE* value (0.008) among all textures, with silt loam having the second smallest median (0.017). Samples of silt loam texture were expected to have the smallest *RMSE* values because silt loam is the dominant texture represented

in the models. Removing the three outlying points of the silt loam distribution decreased the median only slightly, from 0.017 to 0.015. The range of *RMSE* values for sands compared to loams was quite similar, indicating lack of model bias toward a particular texture. The same *RMSE* values divided on the basis of source data are shown in figure 13*B*. The *RMSE* distribution for the Winfield (2003) data set had the tightest range (0.008 to 0.019), with a median value only slightly smaller than the other distributions. This range likely resulted from all 10 samples having silt loam textures. The McElroy and Hubbell (1990) data set displayed the largest range of *RMSE* values (0.001 to 0.054). The θ_{sat} values for five samples were either greatly over- or underestimated, with absolute residual values of 0.07 or more. Over- or underestimation of θ_{sat} could cause the *RMSE* values to be inflated due to the large difference between estimated and measured values of θ near saturation. Textures represented by the McElroy and Hubbell
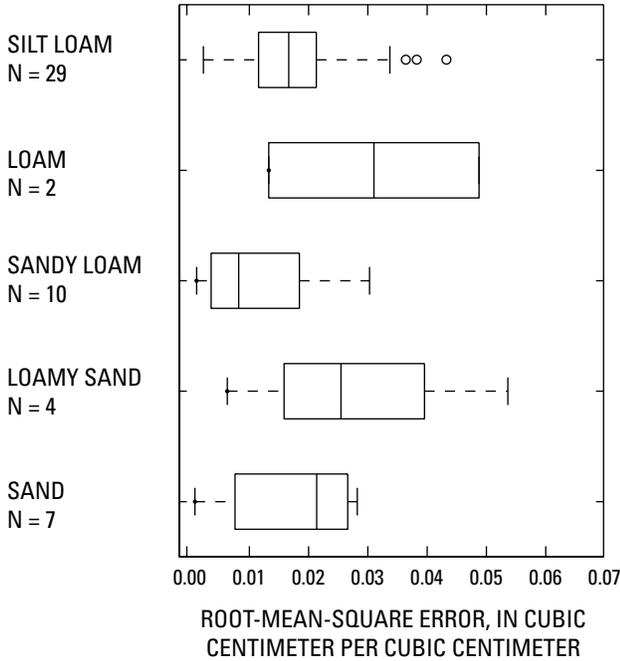
A. Core sample texture
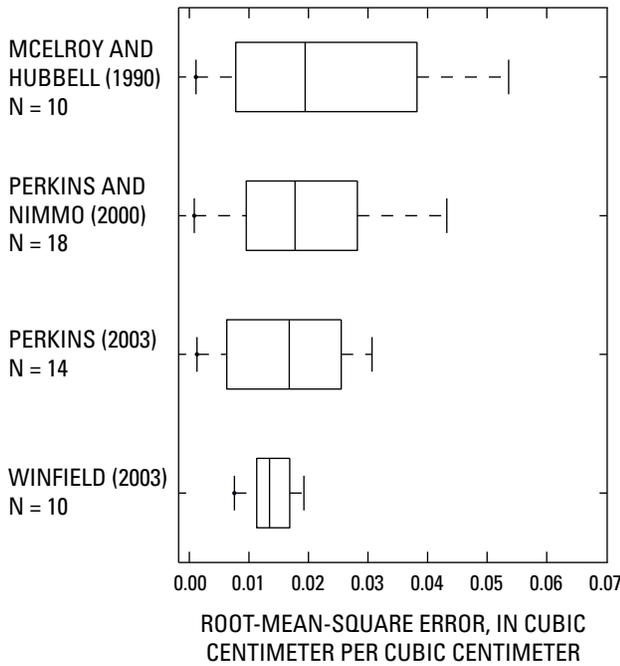
B. Data source for model development

Figure 13.   Root-mean-square-error distributions for the multiple linear-regression model with median particle diameter, uniformity coefficient, and bulk density as input (model combination C-C-C). *RMSE* was calculated for individual core samples between curves estimated from the regression models and fitted to the measured water-retention points. The water-retention curve was represented by the Rossi-Nimmo (1994) junction model.

data set were the most diverse among the four data sets shown in figure 13*B*, with a mixture of silt loams, loams, sandy loams, and loamy sands. This diverse mixture of textures also could explain why the *RMSE* distribution for the McElroy and Hubbell data spanned such a large range.

Comparison of *RMSE* distributions in figure 12 showed that estimated $\theta(\psi)$ curves were insensitive to the way the particle-size distribution was represented; therefore, other criteria were used to select the best particle-size parameter set for the final $\theta(\psi)$ and K_{sat} models. These criteria included: (1) ease of model use or ease of determining the particle-size parameters, (2) best model behavior for each hydraulic parameter (θ_{sat}, ψ_o, λ) in terms of normality and homoscedasticity of the residuals, (3) highest degree of model fit for each hydraulic parameter, based on R^2_{adj}, and (4) best model combination to use for developing a theoretical model from results of the statistical analyses. Model combination A-B-B involved computing two sets of parameters from the particle-size distribution, textural-class percentages and geometric statistics. Although textural-class percentages can be readily determined, calculation of geometric statistics requires good particle-size distribution resolution and points that extend to at least a cumulative frequency of 10 percent on a percent-finer-than basis. Model combination A-B-B was undesirable to use because two sets of particle-size parameters needed to be calculated. Although model combination A-B-B involved regression models with the highest R^2_{adj} values for each $\theta(\psi)$ parameter, the *RMSE* distribution for this model combination differed only slightly from distributions for model combinations A-A-A, B-B-B, and C-C-C. For these reasons, model combination A-B-B was eliminated from further consideration.

Textural-class percentages are the easiest particle-size parameters to determine, followed by the graphical statistics, and then the geometric statistics. The minimum number of particle-size points needed to calculate textural classes is three, chosen to correspond to the textural-class limits of a particular sedimentary or soil classification system. More points are typically measured for those methods where precise determination of particle sizes corresponding to classification limits is difficult (for example, the pipette or hydrometer method). Percent gravel, sand, and silt can be calculated directly from the cumulative particle-size distribution, graphically or by interpolation between measured points. Percent clay is then calculated from the sum of the other three percentages subtracted from 100 percent. The second easiest set of parameters to calculate includes d_{10}, d_{50}, and d_{60}, which can be determined graphically from the cumulative particle-size distribution or calculated by interpolating between measured points. Calculating the graphical statistics requires better particle-size distribution resolution than for calculating textural-class percentages, primarily because the

values of d_{10}, d_{50}, and d_{60} are not defined by specific particle diameters but by cumulative percentiles. Using geometric statistics is attractive because particle-size distributions of sediments often are lognormally distributed (Krumbein, 1938; Pettijohn, 1975). This means that particle-size distribution can be characterized using only two parameters, the mean and standard deviation. However, calculating these parameters requires good resolution of the particle-size distribution. The number of particle-size bins reported for samples in this study varied from 20 for 67 samples (McElroy and Hubbell, 1990; S. Magnuson, written commun., 2002) to 116 for 42 samples (Perkins and Nimmo, 2000; Perkins, 2003; and Winfield, 2003). Geometric statistics can be calculated directly from particle-size data using the moment method (which is sensitive to bin size), or by fitting a lognormal function to the data (which involves access to software with nonlinear regression capabilities). M_g and d_{50} are the two most common measures of central tendency used in the earth sciences. Because a small number of observations can strongly influence M_g, it is not considered a resistant measure of central tendency. However, d_{50} is determined solely by the relative order of the observations in the distribution. The resistant analogs to the geometric statistical parameters are d_{50} and C_u (d_{60}/d_{10}), which have the advantage of not including effects due to skewness or outliers.

To estimate $K(\theta)$, K_{sat} can be combined with $\theta(\psi)$, represented by the Rossi-Nimmo (1994) or van Genuchten (1980) equations, in Mualem's (1976) equation. For ease of application of this step, the K_{sat} and $\theta(\psi)$ PTMs needed to have the same input parameters, which would require calculating only one set of particle-size parameters. For K_{sat}, model A had the highest R^2_{adj} (0.631), with models B and C differing only by 2.5 and 4.2 percent, respectively, from model A. Because the goodness-of-fit values were similar for models A, B, and C for K_{sat}, the final K_{sat} PTM was selected based on the best model combination for $\theta(\psi)$.

Besides ease of use, other criteria used to discriminate between competing models included the best behavior of residuals for each hydraulic parameter (θ_{sat}, ψ_o, λ), the highest R^2_{adj}, and the best parameters to use for development of a theoretical PTM. In general, model combination B-B-B ranked the highest for all of the criteria for model selection except ease of use. Model combination C-C-C ranked second best. Values of R^2_{adj} for each hydraulic parameter were, on average, higher for model B (table 14) than for the other models. Model B also tended to have the most normally distributed and homoscedastic residuals for $\log(\lambda)$ and θ_{sat}, with the normality being only slightly degraded compared to model A for $\log(\psi_o)$. Model combination B-B-B also was the best candidate for use in developing a theoretical model because M_g and σ_g can be related to similar parameters of the pore-size distribution on a lognormal basis (Kosugi, 1996; Hwang and Powers, 2003). Model A involved four explanatory variables that described the particle-size distribution, but because these summed to 100 percent, only three needed to be measured. Model A ranked highest for ease of use because textural data typically are the minimum data reported when describing particle size. However, due to multicollinearity among the textural-class parameters, model A was less desirable to use as the final PTM than models B or C. The parameters for model C readily can be calculated from the cumulative particle-size distribution as long as enough points are available to interpolate accurately d_{10}, d_{50}, and d_{60}. Accurate calculation of model B particle-size parameters depends on particle-size distribution resolution. Resolution affects the width of the bin sizes, and wider bins create larger errors in M_g and σ_g. In addition, distribution's fine end needs to be adequately defined (at least down to a cumulative frequency of 10 percent), otherwise calculated M_g and σ_g values are greatly affected. Because d_{50} and C_u are closely related to M_g and σ_g, respectively, and given that model C ranked second highest for the model criteria, model C was chosen as the final model. The following are the final $\theta(\psi)$ equations using the graphical particle-size statistics and ρ_{bulk} as input:

$$\theta_{sat} = 1.0063 - 0.3998\rho_{bulk}$$
$$-0.0123\log(d_{50}) - 0.0029\log(C_u)$$

$$\log(\psi_o) = -1.4080 + 1.5344\rho_{bulk}$$
$$-0.8394\log(d_{50}) - 0.1510\log(C_u) \quad (12)$$

$$\log(\lambda) = -0.0411 + 0.0974\rho_{bulk}$$
$$+0.1925\log(d_{50}) - 0.4910\log(C_u)$$

The K_{sat} model equation is:

$$\log(K_{sat}) = -1.7690 + 0.0794\rho_{bulk}$$
$$+1.7507\log(d_{50}) - 0.3274\log(C_u) \quad (13)$$

For comparison, the coefficients for the explanatory variables of models A and B are listed in tables 12 through 15 for θ_{sat}, ψ_o, λ, and K_{sat}.

Property-Transfer Model Application at the INEEL

Although multiple-linear regression models are empirical, with results depending on data types and quality used in calibration rather than on physical relationships or processes, their nature does not preclude their usefulness, but may limit their universal applicability. A model that is capable of estimating measurements over great depths and large areas where none were previously available, even if the results are not optimal, is better than having no model at all. Statistical model application is constrained to the data types used in calibration. Because the core sample textures included in the regression models ranged from sands to clays, with most samples classified as silt loams, $\theta(\psi)$ prediction for sediment textures not used in calibration, such as clays or samples with more gravel, may be poor. Measurements of hydraulic and

bulk-physical properties on core samples with more diverse textures are desirable. Development of a PTM that relies more on physical than statistical relationships will be useful because it is not possible to characterize every point in space. A theoretical model developed from the statistical models could be more general in its application to other areas of the INEEL or other geographic locations. Because the calibration data used in this study represent only a small fraction of the total population of sediment properties, the PTMs can only be generalized to the southwestern part of the INEEL until further testing has been completed using samples from other locations at the INEEL. Examining variation in sediment texture and thickness across the INEEL may provide another way to determine locations where the PTMs may be applied.

ρ_{bulk} was the most important bulk-physical property for developing the θ_{sat} model. Inclusion of ρ_{bulk} as an explanatory variable in the PTMs requires continued collection of intact core samples. However, it may be possible to use average ρ_{bulk} values for each textural class or remote sensing of ρ_{bulk} (using neutron, gamma-gamma, or acoustic-velocity logs) in place of core sample collection. Care must be taken in applying remote-sensing techniques because the methods respond differently in the presence of certain minerals, bound water, and high salinity water (Keys, 1988). Additionally, methods that provide total porosity estimates rather than effective porosity would be desirable to use because the ρ_{bulk} values were determined using the total volumes of the core samples. Total volume includes pores that may not be interconnected. Measurements of ρ_{part} could be obtained from borehole cuttings and used with estimates of total porosity to calculate ρ_{bulk} values. Further testing is needed to improve applicability and ease of use of the PTMs presented in this report.

The moderate goodness-of-fit values observed for the θ_{sat}, ψ_o, λ, and K_{sat} models, as measured by the R^2_{adj} values, might indicate missing information in the regression variates. Including other bulk-physical-property data, such as mineralogy or specific surface area, in the regression variates could provide better predictions of θ_{sat}, ψ_o, λ, and K_{sat} than the use of ρ_{bulk} and particle-size data alone. Future adjustments to the PTMs may need to incorporate more diverse types of bulk-physical-property data. As more information is added to the PTMs, the model fits will likely improve. The moderate R^2_{adj} values also could be due to significant errors in the hydraulic and bulk-physical-property measurements. Some experimental methods may have more uncertainty associated with the measurements than other methods, or sparse data may increase the uncertainty of calculations made from the data. Data limitations that could potentially affect the model results include sparseness of particle-size distributions, erroneous measurements (such as θ_{sat} exceeding total porosity), sparseness of $\theta(\psi)$ measurements, and repacking samples prior to hydraulic property measurements. For some core samples used in calibrating the regression models, errors in θ_{sat} and particle size were observed. The approach used in dealing with anomalous measurements was to remove the core samples

from the regression analyses. However, weighted least-squares regression could be used as an alternative approach, where more weight could be given to samples judged higher quality than others.

Most widely-used unsaturated flow and transport models use the van Genuchten (1980) model rather than the Rossi-Nimmo (1994) junction model to represent $\theta(\psi)$ measurements for a particular layer or texture. A conversion between the Rossi-Nimmo (1994) and van Genucthen (1980) parameters is presented in <u>figure 14</u>. The van Genuchten equation is parameterized by θ_{sat}, θ_r, h_o, and N, where the scaling parameter for ψ is h_o (analogous to ψ_o) and the curve-shape parameter is N (analogous to λ). In this study θ_{sat} represents a physical measurement rather than an optimized parameter, and because of this, its regression equation can be used directly in the van Genuchten model. For the 58 potential samples used to calibrate the $\theta(\psi)$ regression models, the relation between $\log(h_o)$ and $\log(\psi_o)$ is shown in <u>figure 14*A*</u> and that between N and λ is shown in <u>figure 14*B*</u>. The R^2 was 0.911 and 0.837, respectively. Setting θ_r equal to zero in the van Genuchten equation and assuming $M = 1 - 1/N$ allows conversion between the Rossi-Nimmo (1994) junction and van Genuchten (1980) model parameters. An alternative way of converting between functional forms of the $\theta(\psi)$ curve involves estimating θ at several ψ values using the Rossi-Nimmo (1994) junction model and the regression equations for each $\theta(\psi)$ parameter, and then fitting the van Genuchten (1980) model (or any other curve-fit model) to these estimated $\theta(\psi)$ points.

Development of a PTM for $K(\theta)$ is needed because both $\theta(\psi)$ and $K(\theta)$ are required to describe unsaturated flow using Richards' equation. Mualem's (1976) model is most commonly used for estimating $K(\theta)$, in combination with van Genuchten's (1980) $\theta(\psi)$ equation. A PTM for $K(\theta)$ can be created by (1) developing regression equations from the existing $K(\theta)$ measurements for 40 core samples (Perkins and Nimmo, 2000; Perkins, 2003; and Winfield, 2003), or (2) combining regression models developed in this study for K_{sat} and $\theta(\psi)$ with the $K(\theta)$ equation of Mualem (1976), or other K models in the literature (Gardner, 1958; Campbell, 1974). Andraski and Jacobson (2000), Andraski (1996), Rossi and Nimmo (1994), and Fayer and others (1992) discuss how to combine the Rossi-Nimmo (1994) junction model with Mualem's (1976) equation. One unsaturated K measurement is better to use as the scaling parameter for K than K_{sat} in Mualem's (1976) equation. K_{sat} includes the influence of macropores, which drain first during desaturation. The shape of the $K(\theta)$ curve does not depend on drainage from the largest pores, but from the pores of the sediment matrix. Using K_{sat} as a scaling parameter for K can give a wrong shape to the $K(\theta)$ curve if macropores are present. Therefore, the first approach, developing regression equations from parameters fit to $K(\theta)$, may be more desirable than an approach that uses K_{sat} as the scaling parameter.

$y = 1.2039x - 0.1220$
$R^2 = 0.911$

LOG(VAN GENUCHTEN MODEL SCALING PARAMETER, IN CENTIMETERS OF WATER)

LOG(ROSSI-NIMMO JUNCTION MODEL SCALING PARAMETER, IN CENTIMETERS OF WATER)

A. Scaling parameter for matric pressure

$y = 0.2147x + 0.2599$
$R^2 = 0.837$

LOG(VAN GENUCHTEN MODEL CURVE-SHAPE PARAMETER)
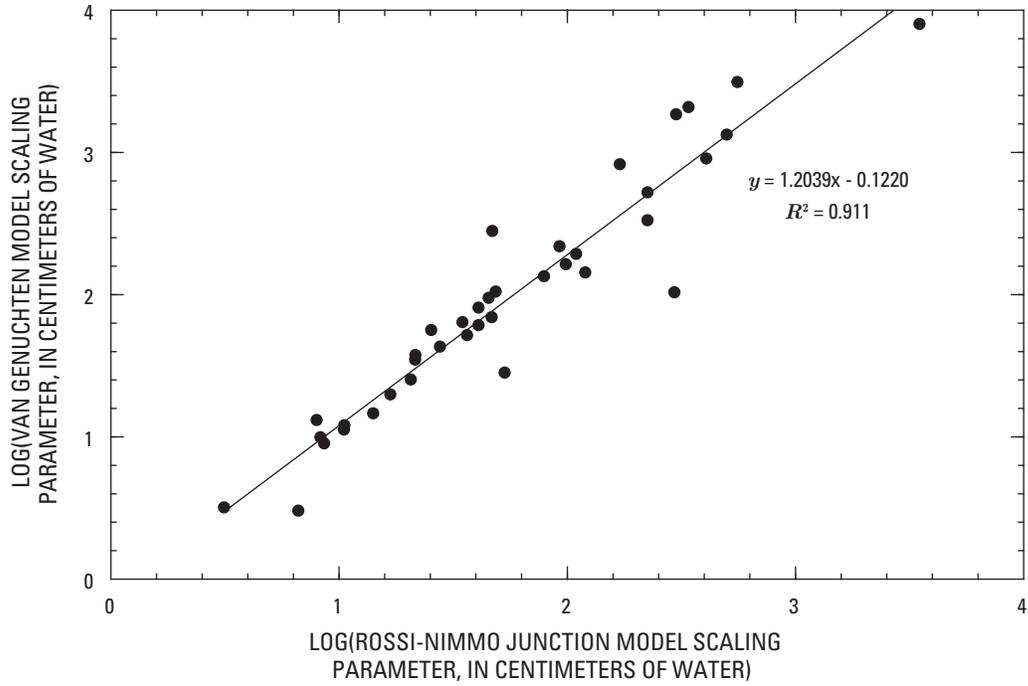
LOG(ROSSI-NIMMO JUNCTION MODEL CURVE-SHAPE PARAMETER)
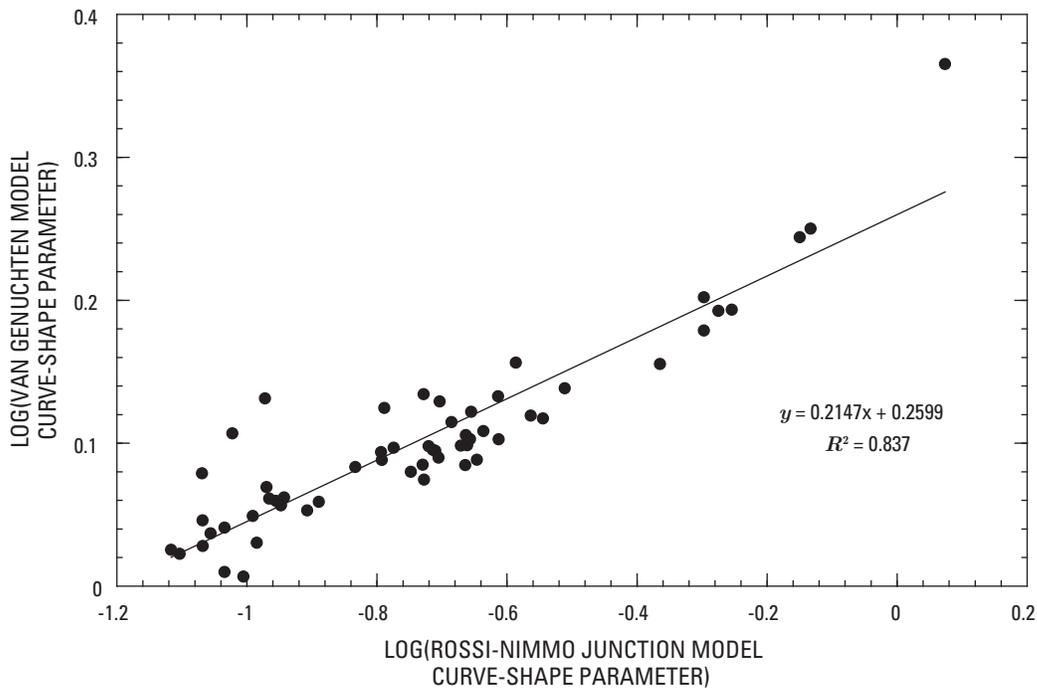
B. Curve-shape parameter

Figure 14.   Relation between water-retention parameters of the Rossi-Nimmo (1994) junction model and the van Genuchten (1980) model.

Testing of the regression models in this report is needed to determine whether the results can be generalized to a wide population of sediments. Model testing can be accomplished by (1) comparing regression models in this study to existing PTMs, (2) completing a new set of core sample measurements, and (3) splitting the original set of measurements into two parts, one for development of the regression equations and one for testing. Other available data sets (Borghese, 1991; Barraclough and others, 1976) and future data sets can be used to test the PTMs developed in this report for $\theta(\psi)$ and K_{sat}, or to adjust the calibration of the models. New measurements can be used to validate the PTMs in this report by comparing the predictions of θ_{sat}, ψ_o, λ, and K_{sat} using equations 12 and 13 to the observed values. Additionally, the predictions can be compared to results from other PTMs not calibrated for the INEEL site to evaluate the need for continued PTM development and the accuracy of the site-specific PTMs. A second set of measurements also could be used to develop separate PTMs that could be compared to the original PTMs in terms of predictive accuracy, and the magnitude and sign of the coefficients. These steps will help ensure that the original PTMs are appropriate for a diverse range of sediment textures and for multiple locations throughout the INEEL. Another way of testing the PTMs involves subdividing the original set of measurements to create one set of observations for model development and one for testing. The limiting factor in this approach, at least for the $\theta(\psi)$ models, is the small number of measurements available. Splitting 58 or fewer observations into two parts leaves only a small data set to work with, which may cause problems in meeting the assumptions of multiple linear regression.

Summary and Conclusions

In this study, multiple-linear regression models were developed for sedimentary interbeds near two facilities at the Idaho National Engineering and Environmental Laboratory, the Radioactive Waste Management Complex and the Vadose Zone Research Park, using available core sample measurements. Separate regression equations were developed for parameters defining the water-retention curve ($\theta(\psi)$) and for saturated hydraulic conductivity (K_{sat}). Three representations of the particle-size distribution were considered as possible input to these models, in addition to bulk density (ρ_{bulk}) and particle density (ρ_{part}). These representations involved the calculation of textural-class percentages (gravel, sand, silt, and clay), geometric statistics (mean and standard deviation), and graphical statistics

(median and uniformity coefficient) from the particle-size data. The Rossi-Nimmo junction model was used to represent the $\theta(\psi)$ measurements. Three parameters define this model: (1) saturated water content (θ_{sat}), (2) a scaling parameter for matric pressure (ψ_o), and (3) a curve-shape parameter (λ). The bulk-physical-property data and optimized hydraulic parameter values were used to develop separate regression models for θ_{sat}, ψ_o, and λ. The predicted parameters then were used to calculate the $\theta(\psi)$ curve from saturation to oven dryness.

The individual regression models for θ_{sat}, ψ_o, λ, and K_{sat} were evaluated based on goodness-of-fit and model behavior, in terms of meeting the assumptions of multiple linear regression. Because ρ_{part} was poorly correlated with θ_{sat}, ψ_o, λ, and K_{sat}, and affected the R^2_{adj} values only marginally, it was not included as an explanatory variable in the regression analyses. The model fits, assessed using the adjusted coefficient of determination (R^2_{adj}), ranged from 0.491 to 0.568 for θ_{sat}, from 0.667 to 0.727 for ψ_o, and from 0.192 to 0.667 for λ. The R^2_{adj} values for K_{sat} were moderate, ranging from 0.589 to 0.632. For θ_{sat} and K_{sat}, the highest R^2_{adj} values were achieved using textural-class percentages and ρ_{bulk} as input. The highest R^2_{adj} for ψ_o and λ were obtained using geometric particle-size statistics and ρ_{bulk} as input. Models involving geometric particle-size parameters tended to have better model behavior, in terms of residual normality and homoscedasticity, than those involving textural-class percentages or graphical particle-size statistics. The geometric statistics are appealing to use because they describe the breadth of particle sizes and average particle size, which may be related directly to similar features of the pore-size distribution. The best models for θ_{sat} and K_{sat} used textural-class percentages and ρ_{bulk} as input, and the best models for ψ_o and λ used geometric statistics and ρ_{bulk} as input. The sensitivity of the model results to the form of the particle-size distribution used as input was tested by examining the combined effect of the θ_{sat}, ψ_o, and λ regression models in predicting the entire $\theta(\psi)$ curve.

The root-mean-square-error (*RMSE*) distributions for four different model combinations were compared, involving: (1) highest R^2_{adj} for θ_{sat}, ψ_o, and λ, (2) textural-class percentages, (3) geometric statistics, and (4) graphical statistics. Because the goodness-of-fit values for the K_{sat} models were similar for each of the three sets of particle-size parameters, the final K_{sat} model was selected based on the best model combination to use for estimating $\theta(\psi)$. For each core sample, the $\theta(\psi)$ curve was estimated using the predicted values of θ_{sat}, ψ_o, and λ from the individual regression equations in the Rossi-Nimmo junction model. This estimated curve was compared to the curve fit to the measured $\theta(\psi)$ points using *RMSE* as a goodness-of-fit indicator. These steps were completed for

each of the four model combinations. *RMSE* distributions then were compared for each competing model to select the best particle-size parameter set to implement. *RMSE* values, in units of volumetric θ, ranged from 0.001 to 0.064 among the four distributions. The *RMSE* distribution for the model combination involving the highest R^2_{adj} values for θ_{sat}, ψ_o, and λ was similar to the distributions that used geometric or graphical statistics as input. Median values for the four distributions were quite similar, ranging from 0.015 to 0.021. The *RMSE* distribution with textural-class percentages used to estimate θ_{sat}, ψ_o, and λ had a slightly larger median than the other combinations, likely due to the low R^2_{adj} of 0.192 for the λ regression model. The *RMSE* distribution based on graphical statistics was further subdivided according to core sample texture and data source for model development. The range of *RMSE* values for sands compared to loams was quite similar, indicating lack of model bias toward a particular texture. Variations in *RMSE* distributions among the source data sets were found to be mainly due to the textures of the core samples analyzed and the over- or underestimation of θ near saturation. From comparisons of *RMSE* distributions among the four model combinations, the predicted θ(ψ) curves were insensitive to the particle-size distribution form used as input to the regression equations.

Given the similarity of the results for each model combination, other criteria were used for final model selection, including: (1) ease of model use or ease of determining the particle-size parameters, (2) optimal regression model behavior and fit, and (3) ability to develop a theoretical model from the statistical results. Ease of model use was deemed the most important criterion for selecting the form of the particle-size distribution to use in the θ(ψ) and K_{sat} models. Although textural-class percentages are the easiest particle-size parameters to determine, multicollinearity among the textural-class parameters limits the applicability of this model. Even though the models with geometric statistics as input generally performed better than the other two particle-size models, geometric statistics are more difficult to calculate and require higher resolution of particle-size distribution than textural-class percentages or graphical statistics. Median particle diameter and uniformity coefficient, which can be calculated directly from a cumulative particle-size distribution, ranked second for ease of determination. The final regression equations for θ_{sat}, ψ_o, λ, and K_{sat} involved ρ_{bulk}, the median particle diameter, and the uniformity coefficient as explanatory variables.

Moderate R^2_{adj} values for θ_{sat}, ψ_o, λ, and K_{sat} may indicate that the bulk-physical-property data used in calibrating the regression models are not sufficient to predict completely the hydraulic parameters or may indicate significant measurement errors in the response or explanatory variables. Other bulk-physical-property data not available for calibrating the property-transfer functions, such as mineralogy or specific surface area, might correlate more strongly with the hydraulic parameters, and may be useful in future regression analyses.

The regression models in this report represent the first attempt to develop a site-specific model for estimating the unsaturated hydraulic properties of sediments at the INEEL. Because the models are the first of their kind for the INEEL, adjustments or recalibration likely will be required in the future as new data become available. These models allow easy determination of hydraulic properties without need for direct measurement. Although collection of core samples still likely will be needed to estimate ρ_{bulk} and particle size, these measurements are less time-intensive to complete than hydraulic-property measurements. The regression models will be useful in estimating hydraulic properties of sediments over large areas, especially as input to unsaturated flow and contaminant transport models. The property-transfer models from this study provide a basis for development of a theoretical model that relies on physical relationships between the pore-size distribution and the bulk-physical properties of the media. A theoretical model based on physical relationships should be more universal in its application throughout the INEEL and other geographic locations.

References Cited

American Society for Testing and Materials (ASTM), 1989, Part 19—soil and rock; building stones: Annual Book of ASTM Standards: Philadelphia, Pa., American Society for Testing and Materials, [variously paged].

Anderson, S.R., 1991, Stratigraphy of the unsaturated zone and uppermost part of the Snake River Plain aquifer at the Idaho Chemical Processing Plant and Test Reactors Area, Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Water-Resources Investigations Report 91–4010 (DOE/ID–22095), 71 p.

Anderson, S.R., and Lewis, B.D., 1989, Stratigraphy of the unsaturated zone at the Radioactive Waste Management Complex, Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Water-Resources Investigations Report 89–4065 (DOE/ID–22080), 54 p.

Anderson, S.R., Liszewski, M.J., and Ackerman, D.J., 1996, Thickness of surficial sediment at and near the Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Open-File Report 96–330 (DOE/ID–22128), 16 p.

Andraski, B.J., 1996, Properties and variability of soil and trench fill at an arid waste-burial site: Soil Science Society of America Journal, v. 60, p. 54-66.

Andraski, B.J., and Jacobson, E.A., 2000, Testing a full-range soil-water retention function in modeling water potential and temperature: Water Resources Research, v. 36, no. 10, p. 3081-3089.

Arya, L.M., and Paris, J.F., 1981, A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data: Soil Science Society of America Journal, v. 45, p. 1023–1030.

Barraclough, J.T., Lewis, B.D., and Jensen, R.G., 1981, Hydrologic conditions at the Idaho National Engineering Laboratory, Idaho—emphasis 1974-1978: U.S. Geological Survey Water-Supply Paper 2191, 52 p.

Barraclough, J.T., Robertson, J.B., and Janzer, V.J., 1976, Hydrology of the Solid Waste Burial Ground as related to the potential migration of radionuclides, Idaho National Engineering Laboratory: U.S. Geological Survey Open-File Report 76–471 (IDO–22056), 183 p.

Barraclough, J.T., Teasdale, W.E., and Jensen, R.G., 1967, Hydrology of the National Reactor Testing Station area, Idaho—annual progress report, 1965: U.S. Geological Survey Open-File Report (IDO–22048), 107 p.

Bartholomay, R.C., 1990, Mineralogical correlation of surficial sediment from area drainages with selected sedimentary interbeds at the Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Water Resources Investigations Report 90–4147, 17 p.

Bartholomay, R.C., 1998, Distribution of selected radiochemical and chemical constituents in perched ground water, Idaho National Engineering Laboratory, Idaho, 1992-95: U.S. Geological Survey Water Resources Investigations Report 98–4026 (DOE/ID–22145), 59 p.

Bartholomay, R.C., Tucker, B.J., Ackerman, D.J., and Liszewski, M.J., 1997, Hydrologic conditions and distribution of selected radiochemical and chemical constituents in water, Snake River Plain aquifer, Idaho National Engineering Laboratory, Idaho, 1992 through 1995: U.S. Geological Survey Water-Resources Investigations Report 97–4086 (DOE/ID–22137), 57 p.

Bechtel BWXT Idaho, LLC, 2003, INEEL Facilities: Idaho Nuclear Technology and Engineering Center—background, accessed March 29, 2004 at http://www.inel.gov/facilities/intec.shtml, last updated March 12, 2003.

Beyer, W.H., ed., 1991, CRC standard mathematical tables and formulae (29th ed.): Boca Raton, Fl., CRC Press, Inc., 609 p.

Blake, G.R., and Hartge, K.H., 1986, Particle density, in Klute, A., ed., Methods of soil analysis, part 1—physical and mineralogical methods (2nd ed.): Soil Science Society of America Book Series No. 5, Madison, Wis., Soil Science Society of America, p. 377-382.

Borghese, J.V., 1991, Hydraulic characteristics of soil cover, Subsurface Disposal Area, Idaho National Engineering Laboratory: Pocatello, University of Idaho, M.S. thesis, 94 p.

Brooks, R.H., and Corey, A.T., 1964, Hydraulic properties of porous media: Colorado State University Hydrology Paper, no. 3, 27 p.

Campbell, G.S., 1974, A simple method for determining unsaturated hydraulic conductivity from moisture retention data: Soil Science, v. 117, p. 311-314.

Campbell, G.S., 1985, Soil physics with BASIC—transport models for soil-plant systems: New York, Elsevier, 150 p.

Campbell, G.S., and Shiozawa, S., 1992, Prediction of hydraulic properties of soils using particle-size distribution and bulk density data, in van Genuchten, M.T., Leij, F.J., and Lund, L.J., eds., International workshop on indirect methods for estimating the hydraulic properties of unsaturated soils, October 11-13, 1989, Proceedings, University of California at Riverside, Calif., p. 317-328.

Cecil, L.D., Orr, B.R., Norton, T., and Anderson, S.R., 1991, Formation of perched ground-water zones and concentrations of selected chemical constituents in water, Idaho National Engineering Laboratory, Idaho, 1986–88: U.S. Geological Survey Water Resources Investigations Report 91–4166 (DOE/ID–22100), 53 p.

Conca, J.L., and Wright, J.V., 1998, The UFA method for rapid, direct measurements of unsaturated transport properties in soil, sediment, and rock: Australian Journal of Soil Research, v. 36, p. 291-315.

Cosby, B.J., Hornberger, G.M., Clapp, R.B., and Ginn, T.R., 1984, A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils: Water Resources Research, v. 20, no. 6, p. 682-690.

Fawcett, R.G., and Collis-George, N., 1967, A filter paper method for determining the moisture characteristics of soil: Australian Journal of Experimental Agriculture and Animal Husbandry, v. 7, no. 24, p. 162-167.

Fayer, M.J., Rockhold, M.L., and Campbell, M.D., 1992, Hydrologic modeling of protective barriers—comparison of field data and simulation results: Soil Science Society of America Journal, v. 56, p. 690-700.

Fetter, C.W., 1994, Applied hydrogeology (3rd ed.): Englewood Cliffs, N.J., Prentice Hall, Inc., 691 p.

Folk, R.L., 1980, Petrology of sedimentary rocks (2nd ed.): Austin, Tex., Hemphill Publishing Company, 184 p.

Gardner, W.R., 1958, Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table: Soil Science, v. 85, p. 228-232.

Greacen, E.L., Walker, G.R., and Cook, P.G., 1987, Evaluation of the filter paper method for measuring soil water suction: International Conference on Measurement of Soil and Plant Water Status, v. 1, p. 137-143.

Gregson, K., Hector, D.J., and McGowan, M., 1987, A one-parameter model for the soil water characteristic: Journal of Soil Science, v. 38, p. 483-486.

Gupta, S.C., and Larson, W.E., 1979, Estimating soil water retention characteristics from particle size distribution, organic matter content, and bulk density: Water Resources Research, v. 15, no. 6, p. 1633-1635.

Hackett, B., Pelton, J., and Brockway, C., 1986, Geohydrologic story of the eastern Snake River Plain and the Idaho National Engineering Laboratory: U.S. Department of Energy, Idaho Operations Office, Idaho National Engineering Laboratory, 32 p.

Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., and Black, W.C., 1998, Multivariate data analysis: Upper Saddle River, N.J., Prentice Hall, Inc., 730 p.

Haverkamp, R., and Parlange, J.Y., 1986, Predicting the water-retention curve from particle-size distribution 1. Sandy soils without organic matter: Soil Science, v. 142, no. 6, p. 325-339.

Haverkamp, R., and Reggiani, P., 2002, Physically based water retention prediction models [property-transfer models], in Dane, J.H., and Topp, G.C., eds., Methods of soil analysis, part 4—physical methods: Soil Science Society of America Book Series No. 5, Madison, Wis., Soil Science Society of America, p. 762-777, 781-782.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Studies in Environmental Science Monograph Series, Amsterdam, Elsevier, v. 49, 522 p.

Hwang, S.I., and Powers, S.E., 2003, Lognormal distribution model for estimating soil water retention curves for sandy soils: Soil Science, v. 168, no. 3, p. 156-166.

Jabro, J.D., 1992, Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data: Transactions of the American Society of Agricultural Engineers, v. 35, no. 2, p. 557-560.

Keys, W.S., 1988, Borehole geophysics applied to ground-water investigations: U.S. Geological Survey Open-File Report 87–539, 305 p.

Kosugi, K., 1996, Lognormal distribution model for unsaturated soil hydraulic properties: Water Resources Research, v. 32, no. 9, p. 2697-2703.

Krumbein, W.C., 1938, Size frequency distributions of sediments and the normal phi curve: Journal of Sedimentary Petrology, v. 8, no. 3, p. 84-90.

Liszewski, M.J., and Mann, L.J., 1992, Purgeable organic compounds in ground water at the Idaho National Engineering Laboratory, Idaho—1990 and 1991: U.S. Geological Survey Open-File Report 92–174 (DOE/ID–22104), 19 p.

Looney, S.W., and Gulledge, T.R., 1985, Use of the correlation coefficient with normal probability plots: The American Statistician, v. 39, p. 75-79.

McElroy, D.L., and Hubbell, J.M., 1990, Hydrologic and physical properties of sediments at the Radioactive Waste Management Complex: EG&G Idaho, Inc., EGG-BG-9147, [variously paged].

Mecke, M., Westman, C.J., Ilvesniemi, H., 2002, Water retention capacity in coarse podzol profiles predicted from measured soil properties: Soil Science Society of America Journal, v. 66, p. 1-11.

Mualem, Y., 1976, A new model for predicting the hydraulic conductivity of unsaturated porous media: Water Resources Research, v. 12, no. 3, p. 513-522.

Nimmo, J.R., 1997, Modeling structural influences on soil water retention: Soil Science Society of America Journal, v. 61, no. 3, p. 712-719.

Nimmo, J.R., and Mello, K.A., 1991, Centrifugal techniques for measuring saturated hydraulic conductivity: Water Resources Research, v. 27, p. 1263-1269.

Nimmo, J.R., Perkins, K.S., and Lewis, A.M., 2002, Steady-state centrifuge [simultaneous determination of water transmission and retention properties], in Dane, J.H., and Topp, G.C., eds., Methods of soil analysis, part 4—physical methods: Soil Science Society of America Book Series No. 5, Madison, Wis., Soil Science Society of America, p. 903-916, 1041–1045.

Nimmo, J.R., Rousseau, J.P., Perkins, K.S., Stollenwerk, K.G., Glynn, P.D., Bartholomay, R.C., and Knobel, L.L., 2004, Hydraulic and geochemical framework of the Idaho National Engineering and Environmental Laboratory vadose zone: Vadose Zone Journal, v. 3, no. 1, p. 6-34.

Nimmo, J.R., Shakofsky, S.M., Kaminsky, J.F., and Lords, G.S., 1999, Laboratory and field hydrologic characterization of the shallow subsurface at an Idaho National Engineering and Environmental Laboratory waste-disposal site: U.S. Geological Survey Water-Resources Investigations Report 99–4263 (DOE/ID–22163), 31 p.

Nimmo, J.R., Stonestrom, D.A., and Akstin, K.C., 1994, The feasibility of recharge rate determinations using the steady-state centrifuge method: Soil Science Society of America Journal, v. 58, p. 49-56.

Perkins, K.S., 2003, Measurement of sedimentary interbed hydraulic properties and their hydrologic influence near the Idaho Nuclear Technology and Engineering Center at the Idaho National Engineering and Environmental Laboratory: U.S. Geological Survey Water-Resources Investigations Report 03–4048 (DOE/ID–22183), 19 p.

Perkins, K.S., and Nimmo, J.R., 2000, Measurement of hydraulic properties of the B-C interbed and their influence on contaminant transport in the unsaturated zone at the Idaho National Engineering and Environmental Laboratory, Idaho: Water-Resources Investigations Report 00–4073 (DOE/ID–22170), 30 p.

Pettijohn, F.J., 1975, Sedimentary rocks (3rd ed.): New York, Harper & Row, Publisher, Inc., 628 p.

Puckett, W.E., Dane, J.H., and Hajek, B.F., 1985, Physical and mineralogical data to determine soil hydraulic properties: Soil Science Society of America Journal, v. 49, p. 831-836.

Rawls, W.J., and Brakensiek, D.L., 1982, Estimating soil water retention from soil properties: Journal of the Irrigation and Drainage Division, American Society of Civil Engineers, v. 108, no. IR^2, p. 166-171.

Reed, M.F., and Bartholomay, R.C., 1994, Mineralogy of selected sedimentary interbeds at or near the Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Open-File Report 94–374, 19 p.

Rightmire, C.T., and Lewis, B.D., 1987, Hydrogeology and geochemistry of the unsaturated zone, Radioactive Waste Management Complex, Idaho National Engineering Laboratory, Idaho: Water Resources Investigations Report 87–4198 (DOE/ID–22073), 89 p.

Ross, P.J., Williams, J., and Bristow, K.L., 1991, Equation for extending water-retention curves to dryness: Soil Science Society of America Journal, v. 55, p. 923-927.

Rossi, C., and Nimmo, J.R., 1994, Modeling of soil water retention from saturation to oven dryness: Water Resources Research, v. 30, no. 3, p. 701-708.

Saxton, K.E., Rawls, W.J., Romberger, J.S., and Papendick, R.I., 1986, Estimating generalized soil-water characteristics from texture: Soil Science Society of America Journal, v. 50, p. 1031-1036.

Schaap, M.G., Leij, F.J., and van Genuchten, M.Th., 1998, Neural network analysis for hierarchical prediction of soil hydraulic properties: Soil Science Society of America Journal, v. 62, no. 4, p. 847-855.

Schaap, M.G., Leij, F.J., and van Genuchten, M.Th., 2001, ROSETTA—a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions: Journal of Hydrology, v. 251, no. 3-4, p. 163-176.

Shakofsky, S.M., 1995, Changes in soil hydraulic properties caused by construction of a simulated waste trench at the Idaho National Engineering Laboratory, Idaho: U.S. Geological Survey Water-Resources Investigations Report 95–4058 (DOE/ID–22121), 26 p.

Soil Survey Staff, 1975, Soil taxonomy—a basic system of soil classification for making and interpreting soil surveys: USDA-SCS Agriculture Handbook no. 436, Washington, D.C., U.S. Government Printing Office, 754 p.

van Genuchten, M.Th., 1980, A closed-form equation for predicting the hydraulic conductivity of unsaturated soils: Soil Science Society of America Journal, v. 44, p. 892-898.

Vereecken, H., Maes, J., Feyen, J., and Darius, P., 1989, Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content: Soil Science, v. 148, no. 6, p. 389-403.

Williams, J., Ross, P., and Bristow, K., 1992, Prediction of the Campbell water retention function from texture, structure, and organic matter, in van Genuchten, M.Th., Leij, F.J., and Lund, L.J., eds., International workshop on indirect methods for estimating the hydraulic properties of unsaturated soils, October 11-13, 1989, Proceedings, University of California at Riverside, Calif., p. 427-441.

Winfield, K.A., 2003, Spatial variability of sedimentary interbed properties near the Idaho Nuclear Technology and Engineering Center at the Idaho National Engineering and Environmental Laboratory, Idaho: Water-Resources Investigations Report 03–4142 (DOE/ID–22187), 36 p.

Wösten, J.H.M., and van Genuchten, M.Th., 1988, Using texture and other soil properties to predict the unsaturated soil hydraulic functions: Soil Science Society of America Journal, v. 52, p. 1762-1770.

Zar, J.H., 1996, Biostatistical analysis (3rd ed.): Upper Saddle River, N.J., Prentice Hall, Inc., [variously paged].

USGS

Development of Property-Transfer Models for Estimating the Hydraulic Properties of Deep Sediments at the Idaho National Engineering and Environmental Laboratory, Idaho