PEER-REVIEWED PUBLICATIONS

<u>Published:</u>

Schaap, M. G., and F. J. Leij. 2000. Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model. *Soil Sci. Soc. Am. J.* 64:843-851.

<u>In Press:</u>

Schaap, M. G., F. J. Leij, and M. Th. Van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol..*

Nemes, A., M.G. Schaap, F.J. Leij and J.H.M. Wösten. 2001. Structure and use of the UNSODA soil hydraulic database V2.0. *J. Hydrol..*

Gonçalves, M. C., F. J. Leij, and M. G. Schaap. 2001. Pedotransfer Functions for Solute Transport Parameters. *European Journal of Soil Science*

Leij, F. J., Ghezzehei, T. A., and Or, D., 2001. Modeling the Dynamics of the Soil Pore-Size Distribution. *Soil and Tillage Research*

<u>Submitted:</u>

Leij, F. J., Ghezzehei, T. A., and Or, D., 2001. Analytical solutions to an approximate model for pore-size evolution after tillage. *Soil Science Society of America Journal.*

# REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.

| 1. AGENCY USE ONLY ( Leave Blank) | 2. REPORT DATE June 12, 2001 | 3. REPORT TYPE AND DATES COVERED Final Report  May 1999-April 2001  21  20 May 01 |
|---|---|---|

| 4. TITLE AND SUBTITLE Development of Pedotransfer Functions with Neural Network Models | 5. FUNDING NUMBERS DAAD19-99-1-0164 |
|---|---|

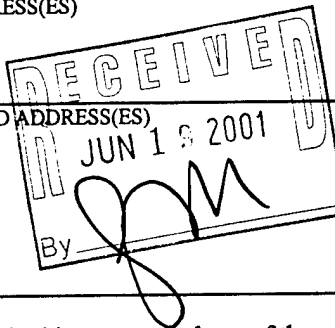6. AUTHOR(S)
F. J. Leij and M. Th. Van Genuchten

| 7. PERFORMING ORGANIZATION NAME(S) AND  ADDRESS(ES) Dept. of Environmental Sciences 2208 Geology Building University of California, Riverside CS 92521 | 8. PERFORMING ORGANIZATION REPORT NUMBER |
|---|---|

| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211 | 10. SPONSORING / MONITORING AGENCY REPORT NUMBER 39153-EV |
|---|---|

RECEIVED
JUN 1 5 2001
By

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

| 12 a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release;  distribution unlimited. | 12 b. DISTRIBUTION CODE |
|---|---|

13. ABSTRACT (Maximum 200 words)
Unsaturated soil hydraulic properties determine the capacity of soils and rocks to retain and transmit water. Hydraulic properties may be needed in applications involving remediation and restoration of contaminated soils, trafficability of soils, flood control, and remotely sensed data. Current methods to measure hydraulic properties are perceived as inadequate to meet the data requirements for most (large scale) applications. Neural networks are used in our research to develop pedotransfer functions (PTFs) for the hierarchical estimation of hydraulic data from basic data such as soil texture and bulk density. Neural networks were calibrated on a database of more than 2000 soils. The predictions generally compared favorably with published PTFs. Especially noteworthy is the unsaturated hydraulic conductivity; we improved its prediction by almost half an order of magnitude compared to traditional methods. We have completed the computer program Rosetta to facilitate neural network based predictions of hydraulic parameters. The uncertainty of the estimates was shown to increase for lower water contents. We have also converted our database of soil hydraulic properties to Windows from DOS.

| 14. SUBJECT TERMS Neural network analysis, pedotransfer functions, software, soil hydraulic properties, vadose zone | | | 15. NUMBER OF PAGES 22 |
|---|---|---|---|
| | | | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED | 18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED | 19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED | 20. LIMITATION OF ABSTRACT UL |
|---|---|---|---|

# DEVELOPMENT OF PEDOTRANSFER FUNCTIONS WITH NEURAL NETWORK MODELS

F. J. Leij[1,2]  and  M. Th. van Genuchten[1]

[1]*George E. Brown Jr. Salinity Laboratory, USDA, ARS, Riverside, CA*
[2]*Dept. of Environmental Sciences, University of California, Riverside, CA*

FINAL REPORT
OF
**ARO Project: 39153-EV**

June 12, 2001

## Table of Contents

**Introduction**

The unsaturated soil hydraulic properties determine the rate at which water and dissolved chemicals move through soils and rocks whose pore space is partially filled with water. These properties (the water retention and hydraulic conductivity curves) govern the rate at which water and dissolved substances move through the vadose zone (upper part) of the soil. Hydraulic properties are increasingly being required in simulation programs used in civil and environmental engineering, soil mechanics, geology, and soil science. Such simulations may lead to a better understanding of flow and transport in the vadose zone and they are therefore increasingly used as a management tool. Among potential Army and civilian applications are: (i) remediation and restoration strategies for sites contaminated through Army actions and the use of military training lands, (ii) trafficability and accessibility of land for different soil and weather conditions, (iii) prediction and control of floods, (iv) interpretation of remotely sensed data, and (v) water flow around man-made objects.

Current methods to measure them are time consuming, expensive and imprecise. They are often perceived as impractical to meet the data requirements for most (large scale) applications. So-called pedotransfer functions (PTFs) are instead used to estimate hydraulic output data from input data such as soil texture, bulk density, organic matter content, and clay mineralogy. All PTFs have a strong degree of empiricism in that they contain model parameters that were calibrated on existing soil hydraulic databases. The PTFs can be as simple as a lookup tables that give hydraulic parameters according to textural class [e.g. *Carsel and Parrish*, 1988; *Wösten et al.*, 1995] or include linear or non-linear regression equations [e.g. *Rawls and Brakensiek*, 1985; *Minasny et al.*, 1999]. There are also PTFs with a more physical foundation, such as the pore-size distribution models by *Burdine* [1953] and *Mualem* [1976], which offer a method to calculate unsaturated hydraulic conductivity from water retention. Models by *Haverkamp and Parlange* [1986] and *Arya and Paris* [1981] use the shape similarity between the particle- and pore-size distributions to predict water retention. *Tyler and Wheatcraft* [1989] combined the Arya model with fractals, while *Arya et al.* [1999a,b] further extended the similarity approach to predict water retention and unsaturated hydraulic conductivity.

The practical application of most PTFs is often hampered by a lack of unsaturated hydraulic data to calibrate and evaluate PTFs. Furthermore, many PTFs have very stringent requirements for the input data. Most authors have usually established PTFs that provided the best results for their data set, often leading to models that require many input variables [cf. *Rawls et al.*, 1991] or models that require very detailed particle size distributions [*Arya and Paris*, 1981; *Haverkamp and Parlange*, 1986]. However, users of PTFs are frequently confronted with situations where one or several input variables needed for a PTF are not available. Another problem is that PTFs provide predictions with a modest level of accuracy and this level is not precisely known It would therefore be useful if PTFs could accept input data with varying degrees of detail and if predictions by PTFs could be accompanied with reliability estimates.

Recently, neural network analysis was used to improve the predictions of empirical PTFs [*Pachepsky et al.*, 1996; *Schaap and Bouten*, 1996; *Minasny et al.*, 1999; *Pachepsky et al.*, 1999]. An advantage of neural networks, as compared to traditional PTFs, is that neural networks require no *a priori* model concept. The optimal, possibly nonlinear, relations that link input data (particle-size data, bulk density, etc.) to output data (hydraulic parameters) are obtained and implemented in an iterative calibration procedure. As a result, neural network models typically extract the maximum information from the data. *Schaap et al.* [1998] used neural network analyses to predict *van Genuchten* [1980] water retention parameters and saturated hydraulic

conductivity. To facilitate practical use of the PTFs, they designed a hierarchical structure to allow input of limited and more extended sets of predictors. The combination with the bootstrap method [*Efron and Tibshirani*, 1993] provided the confidence intervals for the PTF predictions. While neural network-based PTFs may provide relatively accurate predictions, they consist of a large number of coefficients that do not permit easy interpretation or publication in explicit form.

In view of the above, the objectives of this Army-funded project are to: (i) derive PTFs with neural network models and to compare hydraulic properties predicted with these PTFs with those obtained from explicit PTFs already published in the literature, and (ii) develop user-friendly software to implement the neural network models. The first objective was addressed in earlier part of the program period using hydraulic data from the database UNSODA [*Leij et al.*, 1996; *Nemes et al.*, 2001] This is one of the very few public databases that contains unsaturated hydraulic conductivity data. In the latter part of the project period, we developed the Rosetta program for estimating unsaturated hydraulic properties in an effort to meet the second objective.

## Problem Formulation

One-dimensional water flow in a soil or other porous medium that is partially saturated with water is typically described with the mass balance equation according to *Richards* [1931]:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} [ K (\frac{\partial h}{\partial z} - 1)] \tag{1}$$

where $\theta$ is the volumetric soil water content ($L^3 L^{-3}$), $h$ is the soil water pressure head (assumed here to be positive for unsaturated conditions), $K$ is the (unsaturated) hydraulic conductivity ($LT^{-1}$), $t$ is time (T), and $z$ is position (L). The water holding capacity of the soil is determined by the water retention curve, $\theta(h)$, while water transmission is governed by the hydraulic conductivity, $K(h)$ or $K(\theta)$. These functions are highly nonlinear and sometimes hysteretic.

We used the expression of *van Genuchten* [1980] to parameterize retention data:

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = [1 + (\alpha h)^n]^{-m} \tag{2}$$

where the subscripts $r$ and $s$ denote residual and saturated water contents, respectively, $n$ and $m$ are shape factors, whereas $\alpha$ ($L^{-1}$) is inversely related to air entry value. The model for the unsaturated hydraulic conductivity is based on *Mualem* [1976] with $m = 1 - 1/n$:

$$K(S_e) = K_o \, S_e^L [1 - (1 - S_e^{1/m})^m]^2 \tag{3}$$

$$K(h) = K_o \, \frac{\{1 - (\alpha h)^{mn} [1 + (\alpha h)^n]^{-m}\}^2}{[1 + (\alpha h)^n]^{mL}} \tag{4}$$

where $K_o$ is a matching point for the conductivity function at saturation and $L$ is an empirical factor for pore-connectivity and tortuosity. The ability of closed-form expressions to describe unsaturated hydraulic data has been tested by, among others *Leij et al.* [1997]. The above expressions provide an adequate description for most retention and conductivity data.

The number of parameters in the Mualem-van Genuchten model (MVG) for the unsaturated conductivity is frequently reduced by using a measured value for the saturated hydraulic conductivity ($K_s$) as matching point ($K_o$) and by assuming that the pore connectivity and tortuosity factor $L$ can be set equal to 0.5. This approach may be convenient, but may not be very accurate [*Schaap and Leij*, 2000]. Unsaturated flow does not involve macropores thus setting $K_o = K_s$, i.e., using the experimental conductivity at saturation, leads to an overestimation of the unsaturated hydraulic conductivity. Setting $L = 0.5$ is a widely accepted practice, but there

is no physical reason to do so. Rather, this value was selected by Mualem (1976) on empirical grounds. There is no reason why another estimation procedure can not be used. The remaining parameters in Eq.(3) and (4) are essentially retention parameters, which can be estimated more easily and accurately from either retention measurements of PTFs.

Our first task is, with help of the UNSODA database, to evaluate several strategies (PTFs) to predict the unsaturated conductivity function, i.e., $L$ and $K_o$, according to Eq.(3). The major part of the project was devoted to deriving neural network models for the prediction of soil hydraulic properties and to develop a user-friendly program that implements these neural network based PTFs for a hierarchy of input data. The implementation of PTFs based on neural networks is not as straightforward as those based on (linear) regression. In the latter case, the user can predict hydraulic parameters with explicit equations, perhaps a lengthy one, while in the former case a computer algorithm is often necessary for the prediction. The second part of the project was devoted to the development of a user-friendly program called Rosetta, to implement these neural network based PTFs for a hierarchy of input data to predict the output set $\{\theta_r, \theta_s, \alpha, n, K_s, K_o$ and $L)$ to describe water retention, saturated and unsaturated hydraulic conductivity. In the following we will document the program.

## Materials and Methods

### Data Sets

In order to make the PTFs as widely applicable as possible, we obtained a large number of soil hydraulic data and corresponding predictive soil properties from three databases [cf. *Schaap and Leij*, 1998]. The dataset thus assembled contained 2,134 samples for water retention with a total of 20,574 $\theta(h)$ points. Saturated hydraulic conductivity values were available for a subset of 1,306 samples while unsaturated hydraulic conductivity was known for 235 samples with a total of 4,117 $K(h)$ points. The latter subset was solely derived from the database UNSODA [*Leij et al.*, 1996; *Nemes et al.*, 2000] with the requirement that at least five $K(h)$ data points should be available. Figures 1 to 3 give the textural distributions of the datasets for water retention, $K_s$ and unsaturated hydraulic conductivity.

The parameters in Eq.(2) and (3) were fitted to water retention and unsaturated hydraulic conductivity data with the simplex or amoeba algorithm [*Nelder and Mead*, 1965; *Press et al.*, 1988]. The objective function for water retention was:

$$O_W(\mathbf{p}) = \sum_{i=1}^{N_w} (\theta_i - \theta_i')^2 \tag{5}$$

where $\theta_i$ and $\theta_i'$ are the measured and predicted water contents respectively, $N_w$ is the number of measured water retention points for each sample and $\mathbf{p}$ is the parameter vector $\{\theta_r, \theta_s, \alpha, n\}$. For the optimization of unsaturated hydraulic conductivity parameters we used the following objective function:

$$O_K(\mathbf{p}) = \sum_{i=1}^{N_k} [\log_{10}(K_i) - \log_{10}(K_i')]^2 \tag{6}$$

where $K_i$ and $K_i'$ are the measured and predicted hydraulic conductivity respectively, $N_k$ is the number of measured $K(h)$ data points and $\mathbf{p} = \{K_o, L\}$. Logarithmic values of $K_i$ were used in Eq. (6) to avoid a bias towards high conductivities in the 'wet' range. In subsequent analyses, we used log-transformed values of $\alpha, n, K_s$ and $K_o$ to account for their approximately lognormal distributions.

4
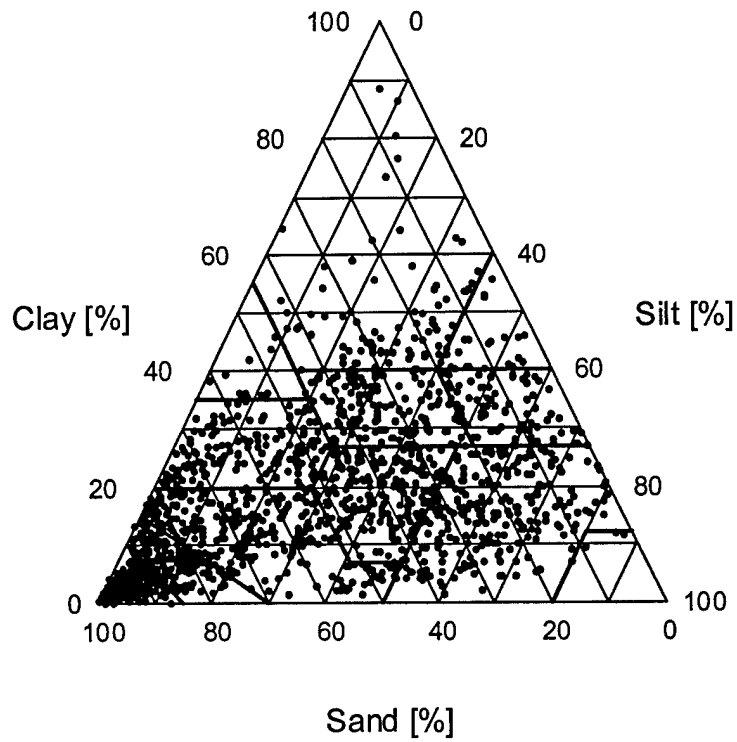
**Figure 1.** Textural distribution of the 2,134 samples for water retention.
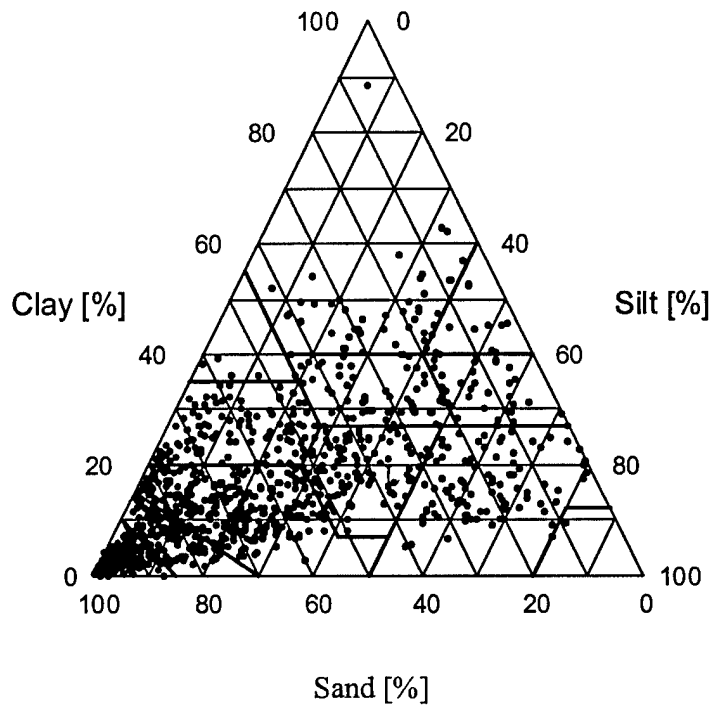


**Figure 2.** Textural distribution of the subset for saturated hydraulic conductivity (1,306 samples).
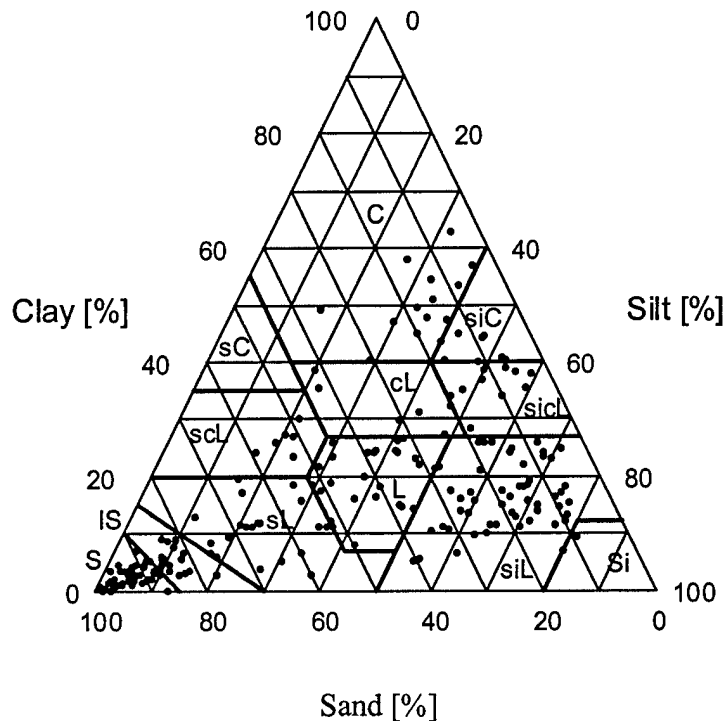
**Figure 3.** Textural distribution of the subset for unsaturated hydraulic conductivity (235 samples). S: sand, lS: loamy sand, sL: sandy loam, scL: sandy clay loam, sC: sandy clay, L: loam, siL: silty loam, Si: silt, sicL: silty clay loam, siC: silty clay, cL: clay loam, C: clay.

*Model Calibration*

Because different numbers of samples were available for water retention, saturated and unsaturated hydraulic conductivity, we developed separate PTFs for each of these characteristics. For the prediction of the water retention parameters ($\theta_r$, $\theta_s$, $\alpha$, n) and $K_s$, we followed a hierarchical approach to estimate hydraulic parameters with limited or more extended sets of input parameters. The first model (H1) consists of a lookup table that that contains parameter averages for each USDA textural class. The second model (H2) uses sand, silt, and clay as input and, in contrast to H1, it provides continuously varying hydraulic parameters across the textural triangle. The third model (H3) includes bulk density to its input while the fourth model (H4) also uses water content at 330 cm pressure. The last model (H5) includes a 15 bar water content in addition to the input variables of the fourth model. The choices of pressure heads were motivated by their availability in the NRCS database [*Soil Survey Staff*, 1995].

The hierarchical approach is uncertain for the unsaturated hydraulic conductivity because prior to his project no reliable data sets existed to investigate the dependency of unsaturated hydraulic conductivity on texture, bulk density and other parameters. *Schaap and Leij* [2000] first investigated the prediction of the unsaturated hydraulic conductivity. The results, which will be given shortly, indicate that $K_0$ and $L$ could be predicted from fitted water retention parameters ($\theta_r$, $\theta_s$, $\alpha$ and n) rather than textural data. For the development of Rosetta, we also investigated how well we can predict $K_0$ and $L$ using *predicted* retention parameters obtained from models H1...H5. This method does not require fitted retention parameters and allows $K_0$ and $L$ to be estimated in a pseudo-hierarchical manner.

6

While model H1 is a simple table with average hydraulic parameters for each textural class, all other models were established using a combination of neural networks and the bootstrap method. Each of these PTFs consist of 60 (water retention) or 100 (saturated and unsaturated conductivity) neural networks models. These neural network models are all slightly different because they were calibrated on different subsets of the original dataset as generated by the Bootstrap Method (Efron and Tibshirani, 1993). For the sake of brevity, we refer to Schaap et al. (1998), Schaap and Leij, (1998), and Schaap and Leij (2000) for more information about the model calibration.

*Analysis*

The performance of the models was evaluated according to three error criteria using the same data that was used for model calibration. We computed the coefficient of determination ($R^2$) between predicted and fitted hydraulic parameters. Further, we computed the root mean square error (RMSE) between measured and predicted water contents, saturated hydraulic conductivity and unsaturated hydraulic conductivity:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\zeta_i' - \zeta_i)^2}$$

In addition, we computed the mean error (ME) to quantify systematic errors with:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(\zeta_i' - \zeta_i) \quad [5]$$

The symbols $\zeta$ and $\zeta'$ denote measured or predicted $\theta(h)$, log $K_s$, or log $K(h)$ values; $N$ is the number of measurements over which the RMSE and ME values were calculated. The RMSE and ME values will be given with w, s, and u subscripts for water retention, saturated, and unsaturated hydraulic conductivity, respectively. Predicted water retention or unsaturated hydraulic conductivity values were calculated by evaluating the hydraulic functions at the pressure heads of the measurements. Because logarithmic values were used for $K_s$ and $K(h)$, the corresponding RMSE and ME values are dimensionless; the units of $RMSE_w$ and $ME_w$ are in $cm^3/cm^3$. In this study, we will compute the RMSE and ME values over all available data (i.e. $N_w$=20,574 for retention, $N_s$ =1,306 for $K_s$, and $N_u$=4,117 for $K(h)$). To investigate how the RMSE and ME values vary with pressure head, we will also compute these values for ten pressure head ranges between 0, 3.2, 10, 32, 100, 320, 1000, 3200, 10000, 32000, and 100000 cm.

**Results**

*Unsaturated Hydraulic Conductivity*

The optimization results of are presented in Table 1 as averages for each textural group or all 235 samples. Because logarithmic values were used for $K_i$, $RMSR_K$ results are dimensionless. For the water retention curve, we found average $RMSE_W$ values between 0.0096 and 0.0141 $cm^3cm^{-3}$. Contrary to our expectation, average $log_{10}(\alpha)$ were higher for the Loams than for the Sands while we also found that the Loams had $log_{10}(n)$ values that were lower than the Clays. For the hydraulic conductivity curve we found $RMSE_K$ values between 0.393 to 0.481, which corresponds to an error of about 0.4 orders of magnitude. The relatively small variation in $RMSE_K$ indicates a good description of log hydraulic conductivity for all textural groups. Average $K_o$ values were almost an order of magnitude lower than average $K_s$. The

**Table 1.** Optimized hydraulic parameters for each soil textural group with standard deviations in parentheses. For $\alpha$, $n$, $K_s$ and $K_o$ the mean values and standard deviations were computed after the $\log_{10}$ tranformations.

| | N | $\theta_r$ cm³cm⁻³ | $\theta_s$ Cm³cm⁻³ | $\log_{10}(\alpha)$ cm⁻¹ | $\log_{10}(n)$ | RMSE w cm³cm⁻³ | $\log_{10}(K_s)$ cm day⁻¹ | $\log_{10}(K_o)$ cm day⁻¹ | L | RMSE K |
|---|---|---|---|---|---|---|---|---|---|---|
| All | 235 | 0.055 (0.073) | 0.442 (0.101) | -1.66 (0.52) | 0.214 (0.209) | 0.0122 | 1.92 (0.81) | 1.03 (1.04) | -3.09 (8.75) | 0.410 |
| Sands[†] | 100 | 0.052 (0.043) | 0.396 (0.056) | -1.58 (0.37) | 0.349 (0.228) | 0.0122 | 2.24 (0.79) | 1.29 (1.06) | -1.28 (3.17) | 0.395 |
| Loams[§] | 41 | 0.056 (0.091) | 0.512 (0.132) | -1.39 (0.50) | 0.076 (0.047) | 0.0119 | 2.03 (0.64) | 1.42 (0.98) | -6.97 (9.57) | 0.398 |
| Silts[¶] | 58 | 0.031 (0.058) | 0.428 (0.078) | -1.92 (0.52) | 0.139 (0.141) | 0.0141 | 1.70 (0.61) | 0.82 (0.80) | -1.22 (10.17) | 0.403 |
| Clays[#] | 36 | 0.098 (0.109) | 0.512 (0.108) | -1.75 (0.64) | 0.114 (0.112) | 0.0096 | 1.31 (0.80) | 0.26 (0.94) | -5.96 (12.40) | 0.481 |

†: Sand, Loamy Sand, Sandy Loam, Sandy Clay Loam.
§: Loam, Clay Loam.
¶: Silty Loam, Silt.
#: Clay, Sandy Clay, Silty Clay, Silty Clay loam.

difference may be caused by soil structure that allows macropore flow at or near saturation [cf. *Luckner et al.* 1989]. Unsaturated flow does not involve macropores thus causing much lower $K_o$ to be inferred from unsaturated hydraulic conductivity. The results in Table 1 indicate that setting $K_o{=}K_s$ , i.e., using the experimental conductivity at saturation, leads to an overestimation of the unsaturated hydraulic conductivity. However, use of a the more accurate fitted rather than the measured value for $K_o$ will lead to additional errors in the conductivity near saturation.

The prediction of $K_o$ and $L$ was investigated for potential predictors like sand and clay percentages, bulk density, $\theta_r$, $\theta_s$, $\alpha$, $n$, and $K_s$. A Spearman rank correlation (e.g. Press et al., 1988) between potential input and output variables $K_o$ and $L$ was determined to select promising input variables. Subsequently, we developed neural network models to predicted $K_o$ and $L$. Three models for predicting $K_o$ and $L$ were tested. Model A reflects the traditional Mualem-van Genuchten model [*van Genuchten*, 1980] with $K_o{=}K_s$ and $L{=}0.5$. In the case of model B, we constructed neural network models that predicted only $K_o$ while assuming that $L{=}{-}1$. Model C predicts both $K_o$ and $L$. Based on the Spearman rank correlation, we used four different sets of predictors for models B and C: 1) sand and clay percentages and bulk density, 2) retention parameters $\{\theta_r, \theta_s, \alpha, n\}$, 3) $\{\theta_r, \theta_s, \alpha, n\}$ and $K_s$, and 4) sets 1 and 3 combined. Models B and C were indexed according to these four sets of predictors (i.e., B1..B4, C1..C4).

Table 2 shows that the $RMSE_K$ of model A is more than one order of magnitude, with a very high value for the Clays (1.70). Models B1 and C1 yielded somewhat lower $RMSE_K$ values by using sand, silt, clay and clay (SSCBD) as predictors. Models B2 and C2 clearly show that water retention parameters make more effective predictors. Results for models B3 and C3 show that adding $K_s$ to the retention parameters increased the coefficients of determination of $K_o$. However, comparison with models B2 and C2 shows that $RMSE_K$ were only marginally reduced, if at all. Models C2, C3, and C4 had lower $RMSE_K$ than the B2, B3, and B4 models, especially for the Loams and Clays. This suggests that both $K_o$ and $L$ should be predicted.

**Table 2.** Coefficients of determination and $RMSE_K$ results for predictions of unsaturated hydraulic conductivity according to models A, B and C

| Model | Input | $R^2$ | | $RMSE_K$ | | | | |
| | | $\log_{10}$ $(K_o)$ | L | All | Sands | Loams | Silts | Clays |
|---|---|---|---|---|---|---|---|---|
| A | $K_o=K_s$, $L=0.5$ | - | - | 1.31 | 1.22 | 1.35 | 1.20 | 1.70 |
| Flexible $K_o$, $L=-1$ | | | | | | | | |
| B1 | SCBD[†] | 0.29 | - | 1.16 | 1.04 | 1.54 | 0.93 | 1.40 |
| B2 | $\theta_r\theta_s\alpha\,n$ | 0.51 | - | 0.95 | 0.90 | 1.30 | 0.73 | 1.08 |
| B3 | $\theta_r\theta_s\alpha\,n\,K_s$ | 0.64 | - | 0.96 | 0.88 | 1.26 | 0.78 | 1.16 |
| B4 | SCBD$\theta_r\theta_s\alpha\,n$ $K_s$ | 0.63 | - | 0.95 | 0.87 | 1.22 | 0.78 | 1.13 |
| Flexible $K_o$ and $L$ | | | | | | | | |
| C1 | SCBD | 0.27 | 0.08 | 1.18 | 1.16 | 1.48 | 0.86 | 1.43 |
| C2 | $\theta_r\theta_s\alpha\,n$ | 0.48 | 0.45 | 0.84 | 0.91 | 0.88 | 0.69 | 0.86 |
| C3 | $\theta_r\theta_s\alpha\,n\,K_s$ | 0.56 | 0.43 | 0.82 | 0.86 | 0.92 | 0.68 | 0.82 |
| C4 | SCBD$\theta_r\theta_s\alpha\,n$ $K_s$ | 0.53 | 0.40 | 0.79 | 0.83 | 0.84 | 0.67 | 0.81 |

† SCBD: sand, clay and bulk density.

Although model C4 performed the 'best' in terms of $RMSE_K$, the differences with models C2 and C3 were relatively small. Given the fact that C2 required only retention parameters to predict $K_o$ and $L$, this model is preferable because retention parameters are already required a-priori to compute $S_e$. It is interesting to note that model C2 uses less data than model A ($K_s$ is not required), yet it has a $RMSR_K$ that is about 0.5 order of magnitude lower. Direct measurement of conductivity leads, of course, to more accurate results than estimation with any of the neural network models. $RMSE_K$ of models C2, C3, and C4 are significantly higher (approximately 0.8) than the $RMSE_K$ of the direct fit to the measured K-h data (Table 1, 0.41). The results for direct fit are essentially averages for *individual* samples and are insensitive to variations among samples due to differences in, for example, measurement methods. The neural network models, however, attempt to implement relations that are valid for *all* the samples in the calibration data set. Consequently, the $RMSE_K$ of the model predictions are sensitive to systematic and random differences that exist from sample to sample.

*Calibration for Rosetta*

An overview of the performance of the hierarchical models for prediction of water retention parameters and $K_s$ is given in Table 3. Not surprisingly, the results show that correlations between fitted and predicted parameters increase and RMSE values decrease when more predictors are used (H1...H5). Residual water content is difficult to predict with any model, while saturated water content is difficult to predict without information about bulk density. The correlation for $\alpha$ increases considerably when one or two retention points are added to the predictors (H4 and H5). The $n$ parameter and $K_s$ generally have the highest correlations showing a gradual increase in $R^2$ from Model H1 to H5.

9

**Table 3.** $R^2$ and RMSE values for the five hierarchical models to predict water retention parameters and saturated hydraulic conductivity. SSC: percentages sand, silt and clay; BD: bulk density; $\theta_{33}$, $\theta_{1500}$: water contents at 330 and 15 000 cm pressure. The $RMSE_w$ for the direct fit to water retention data is also shown.

| Model | Input | Water retention | | | | | Saturated Conductivity | |
|---|---|---|---|---|---|---|---|---|
| | | $R^2$ | | | | $RMSE_w$ | $R^2$ | $RMSE_s$ ($ME_s$) |
| | | $\theta_r$ | $\theta_s$ | Log $\alpha$ | Log $n$ | $cm^3/cm^3$ | Log $K_s$ | (-) |
| H1 | Textural Class | 0.066 | 0.143 | 0.203 | 0.452 | 0.078 | 0.427 | 0.739 (-0.001) |
| H2 | SSC | 0.086 | 0.178 | 0.238 | 0.473 | 0.076 | 0.461 | 0.717 (-0.001) |
| H3 | SSCBD | 0.094 | 0.581 | 0.265 | 0.495 | 0.068 | 0.535 | 0.666 ( 0.000) |
| H4 | SSCBD$\theta_{33}$ | 0.121 | 0.605 | 0.417 | 0.599 | 0.047 | 0.640 | 0.586 (-0.004) |
| H5 | SSCBD$\theta_{33}\theta_{1500}$ | 0.387 | 0.600 | 0.577 | 0.760 | 0.044 | 0.647 | 0.581 (-0.002) |
| | | | | | | | | |
| Direct fit to data | | - | - | - | - | 0.012 | - | - |

Figure 4 shows the $RMSE_w$ for models H1...H5 for water retention for ten pressure head classes as well as the number of water retention points in each class (bars). The $RMSE_w$ of the direct fit to the data (F) serves as the minimum possible error. No PTF can predict water retention better than this fit. Models H1 and H2 show a very similar pattern, indicating that there is not much to be gained by using textural classes or the sand, silt or clay percentages without other predictors. Model H3, however, shows considerable improvement near saturation and a better performance until $h$=3,200 cm (i.e. log $h$=3.5). Usage of a water content at 330 cm (H4) improves the prediction between 10 and 10 000 cm, whereas addition of a water content at 15 bar provides a further improvement beyond 100 cm. Because of the small number of observations, the graph is unreliable for $h$ > 32,000 cm.

Figure 5 shows a similar picture for $ME_w$; negative numbers denote underestimation by the models. Clearly, the direct fit to the data (F) closely adheres the line of $ME_w$=0, indicating that Eq. (2) adequately describes retention data over the entire pressure head range. All models underestimate water retention near saturation ($h$<3.2 cm or log h < 1.5) and overestimate water contents between 3.2 and 10 cm; all models underestimate water retention beyond 32 cm. Models H1...H3 all have a very similar behavior, while models H4 and H5 make smaller systematic errors for $h$ >100 cm. Mean errors for $K_s$ are essentially equal to zero (Table 1).
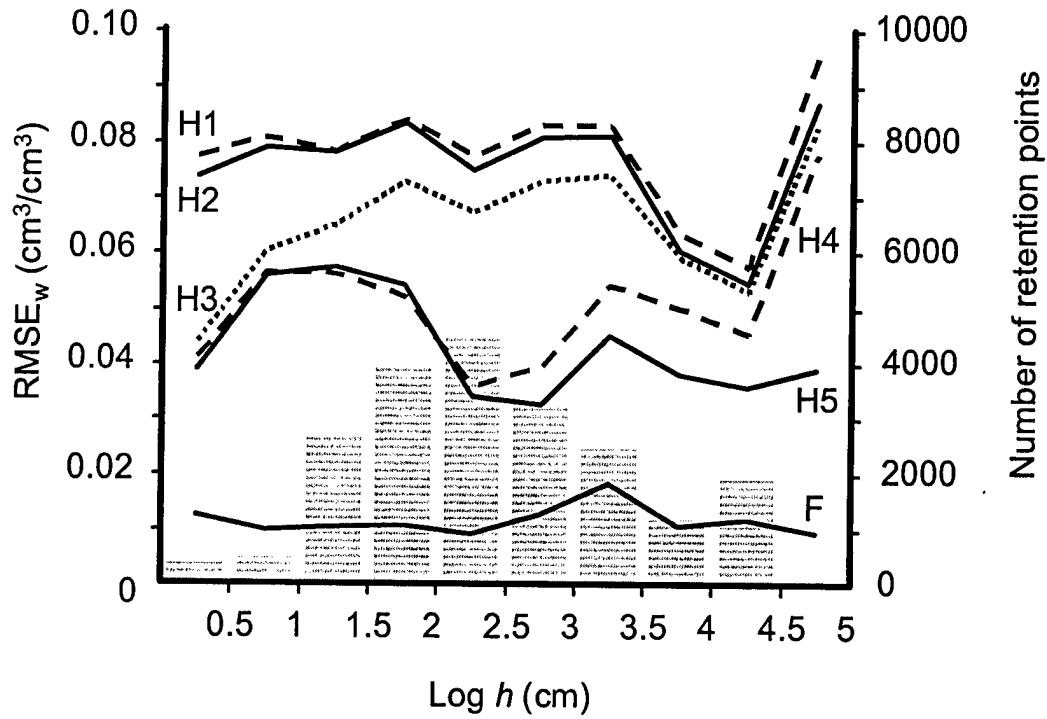
**Figure 4.** $RMSE_w$ (lines, left axis) of the direct fit to water retention data (F) and the five hierarchical models (H1…H5). The number of retention points for each pressure head class is also shown (bars, right axis).
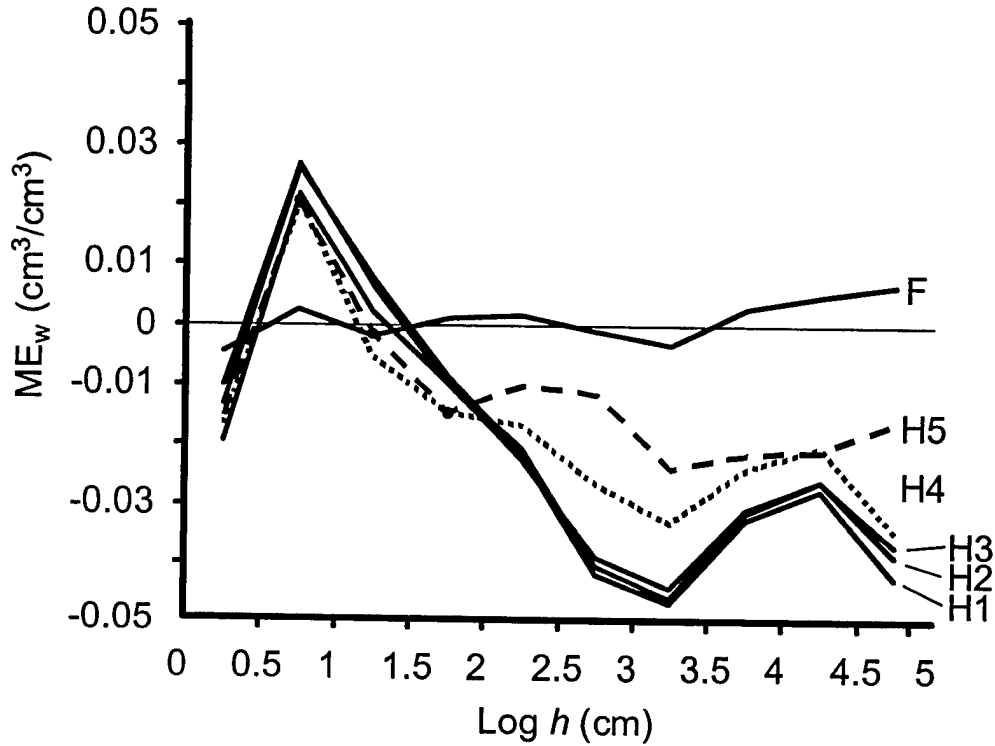


**Figure 5.** $ME_w$ of the direct fit to water retention data (F) and the five hierarchical models (H1…H5). A negative value indicates an underestimation of water contents.

Results for three different methods to predict $K(h)$ appear in Table 4. The first method (MVG) is the traditional Mualem-van Genuchten model ($K_o=K_s$ and $L=0.5$. The second method (C2-Fit) is model C2 of *Schaap and Leij* [2000] that predicts $K_o$ and $L$ from fitted retention parameters. The third method also uses model C2 but with *predicted* retention parameters derived from models H1 to H5 as input (denoted as: C2-H1...C2-H5). In this case, we also use predicted $n$ and $S_e$ values in Eq. (3), which are derived from predicted $\alpha$ and $n$ (Eq. (1)).

MVG model clearly provides the poorest predictions of $K(h)$ with an average RMSE$_u$ of 1.40 (i.e. 1.4 order of magnitude). *Schaap and Leij* [2000] showed that its prediction was especially poor for clayey soils (RMSE$_u$=1.70). RMSE$_u$ increased to about 1.5 for hydraulic parameters predicted with models H1...H5 (data not shown). Model C2-Fit has an RMSE$_u$ that is almost half an order of magnitude lower (0.79), while it also has a more uniform performance over all textural classes (see *Schaap and Leij*, 2000). As expected, models C2-H1...C2-H5 do not perform as well as model C2-Fit, but better than the MVG model. For example, model C2-H5 has an RMSE$_u$ that is only slightly higher than that of C2-Fit (0.90 vs. 0.79). Note however, that the correlations for $K_o$ and $L$ of models C2-H1...C2-H4 are extremely poor and only slightly better for model C2-H5.

Figure 6 shows the RMSE$_u$ for ten pressure head classes as well as the number of conductivity measurements in each class (bars). The direct fit of in Eq. (3) to the data (F) indicates the minimum attainable error for all models with flexible $K_o$ and $L$. We note that the direct fit has a relatively large error near saturation. As expected, model C2-Fit has the best overall performance. Model C2-H5 has a very similar performance between $h=10$ and 3,200 cm ($1 < \log h < 3.5$), while models C2-H1...C2..H4 have similar performances in the range between 10 and 1000 cm. The MVG model has the worst overall performance, except for near saturated conditions where it performs better than all other models – even better than the direct fit of $K_o$ and $L$. This is a result of the assumption that $K_o=K_s$. The RMSE$_u$ of all models strongly increases beyond 3,200 cm. However, in this range there are only a few data points.

Figure 7 shows the mean error for the ten pressure head groups; negative values indicate an underestimation of unsaturated hydraulic conductivity. The MVG model exhibits a strong overestimation of hydraulic conductivity over the entire pressure head range. Models C2-Fit and C2-H1 through C2-H5 are all based on fitted $K_o$ and $L$, and therefore, they show similar problems of underestimating conductivity between 0 and 32 cm. The mean errors are near zero between 32 and 1000 cm, but rapidly become more negative beyond 1000 cm. This is probably caused by an inability to accurately predict the $L$ parameter, which controls the slope of the conductivity curve under dry circumstances. Again, model C2-Fit has the best performance, followed by Model C2-H5.
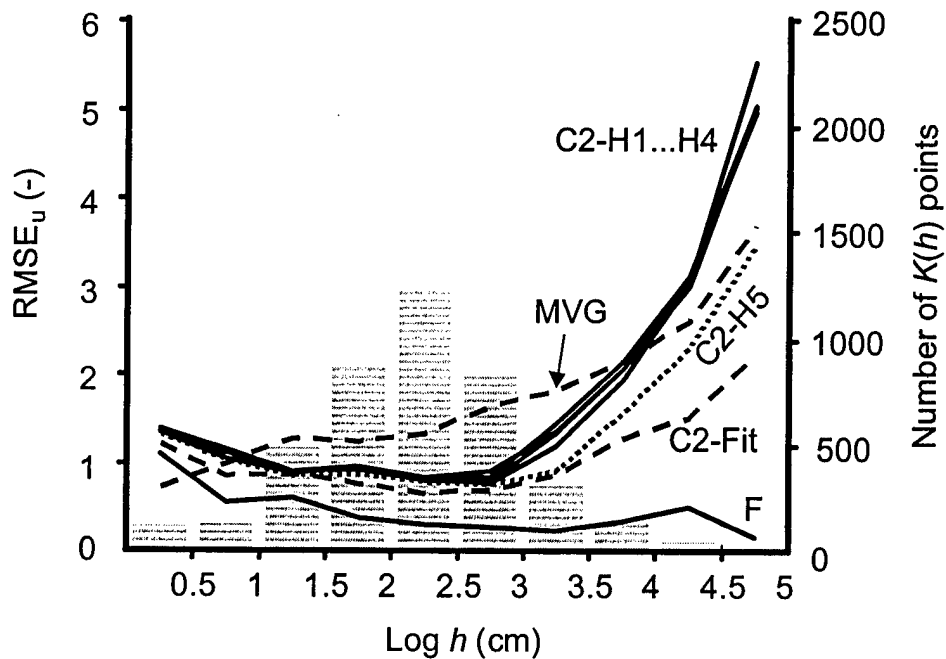
**Figure 6.** RMSE_u (lines, left axis) for the direct fit to conductivity data (F), for model C2 using fitted retention parameters (C2-FIT), model C2 using predicted retention parameters from the hierarchical approach (C2-H1...C2-H5). The Mualem-van Genuchten model with $K_0=K_s$ and $L=0.5$ is shown as MVG. The number of conductivity points per pressure head class is also shown (bars, right axis).
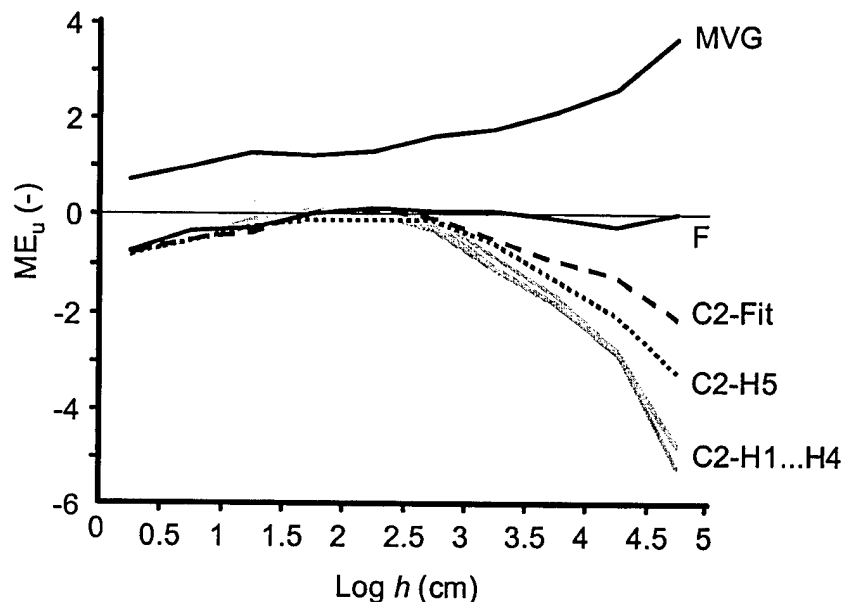


**Figure 7.** ME_u for the direct fit to conductivity data (F), for model C2 using fitted retention parameters (C2-FIT), model C2 using predicted retention parameters from the hierarchical approach (C2-H1...C2-H5). The Mualem-van Genuchten model with $K_0=K_s$ and $L=0.5$ is shown as MVG.

13

## Discussion

The results of our study showed that the traditional use of the Mualem-van Genuchten model to predict the unsaturated conductivity, i.e., with $K_o=K_s$ and $L=0.5$, leads to relatively poor predictions of unsaturated hydraulic conductivity (RMSE$_K$=1.31). A neural network model was derived that predicts both $K_o$ and $L$ from retention parameters ($\theta_r$, $\theta_s$, $\alpha$, $n$) with a RMSE$_K$ of 0.84, which was a substantial improvement over the traditional Mualem-van Genuchten model (cf. Table 2).

The calibration of PTFs for the Rosetta program demonstrated that there is no such thing as a "free lunch" when predicting instead of measuring hydraulic properties. Even with the best models, H5 for retention and C2-Fit for unsaturated conductivity, the correlations between predicted and fitted or measured hydraulic parameters are modest at best (cf. Table 3 and 4). The differences between RMSE and ME values of prediction and direct fits (Fig. 4-7) further suggest that the models in this study could be improved upon. However, the direct fits only give the theoretically minimum attainable errors for PTFs because Eq. (2) and (3) are fitted to *individual* characteristics. Therefore, the fit ignores any effects that cause variation among hydraulic properties, such as variability in physical soil properties or systematic differences among measurement methodologies. In contrast, the PTFs are supposed to be valid for the *ensemble* of all characteristics. This problem is illustrated in Fig. 8 in which we plotted retention data for a narrow selection of 47 loam samples with bulk densities between 1.3 and 1.4 gr/cm$^3$. The average retention curve, as predicted by H3, is also shown. We expect to see retention data in a narrow band, but the figure shows that there is a considerable scatter. This scatter of data points is unlikely to be caused by the narrow range in texture or bulk density. Rather, the variation may be caused by factors other than texture and bulk density or by systematic differences in measurement methodologies.

Improved PTFs were obtained by *Vereecken et al.* [1989] and *Schaap and Bouten* [1996] who used more particle size fractions. Other predictors can also be used to improve the performance of the models, such as: organic matter content, porosity, particle density soil chemical parameters, soil structure, mineralogy, pedality, among others (cf. Rawls et al., 1991). However, using more predictors also requires that they be available for both calibration of PTFs *and* their application. This causes a scenario where measurement of input parameters diminishes the advantages of PTFs over direct measurement of hydraulic properties. From this perspective, the inclusion of measured water retention points [cf. *Ahuja et al.*, 1989; *Messing*, 1989; *Williams et al.*, 1992; *Schaap and Bouten*, 1996] is a pragmatic way to improve the prediction of PTFs. Water retention points can be viewed as lumped parameters that contain implicit information about hydraulic properties not provided by soil texture or bulk density. In many cases, one or two retention points can be measured relatively quickly or are available in national databases such as the NRCS database, which contains more than 120,000 soil horizons for the USA [*Soil Survey Staff*, 1995]. Using the same data as in Fig 8, we plotted predicted versus measured retention points for model H3 and H5 in Fig. 9. While the agreement with the measured data is not perfect, the predictions by model H5 are much better than the predictions by model H3. If retention points are not available, models H1...H3 may still make acceptable predictions – also considering that these models were calibrated on the same data as models H4 and H5.
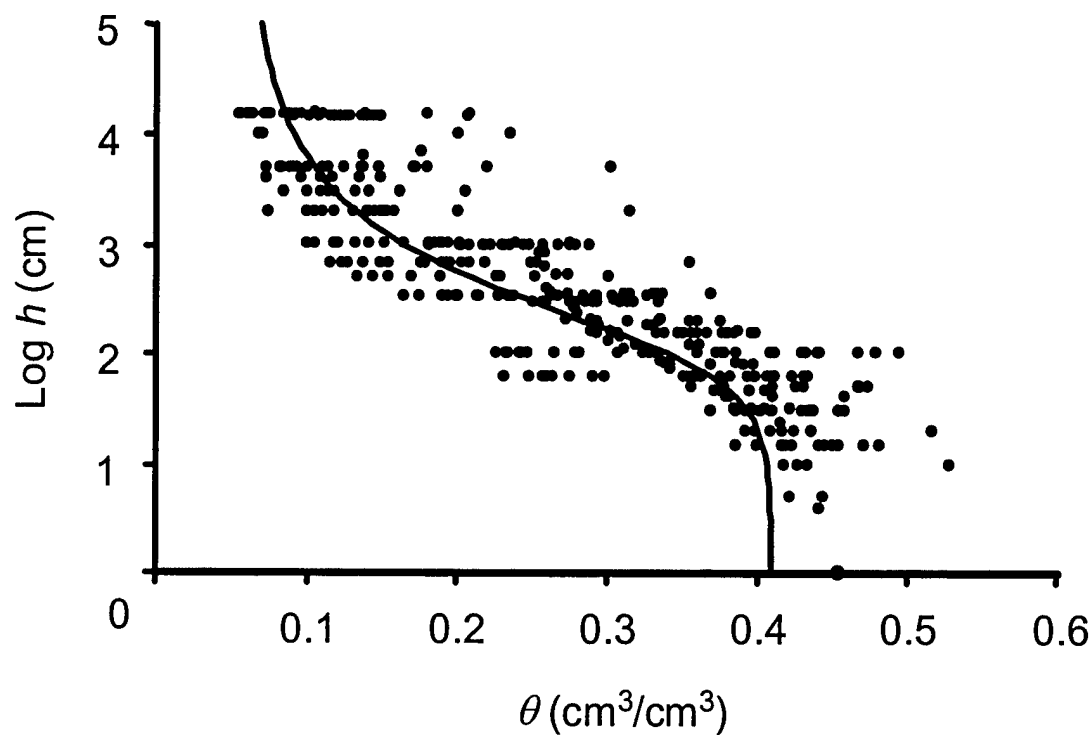
**Figure 8.** Water retention data for 47 loam samples (totaling 412 points) with a bulk density range between 1.3 and 1.4 g/cm$^3$. The curve represents the prediction with model H3.
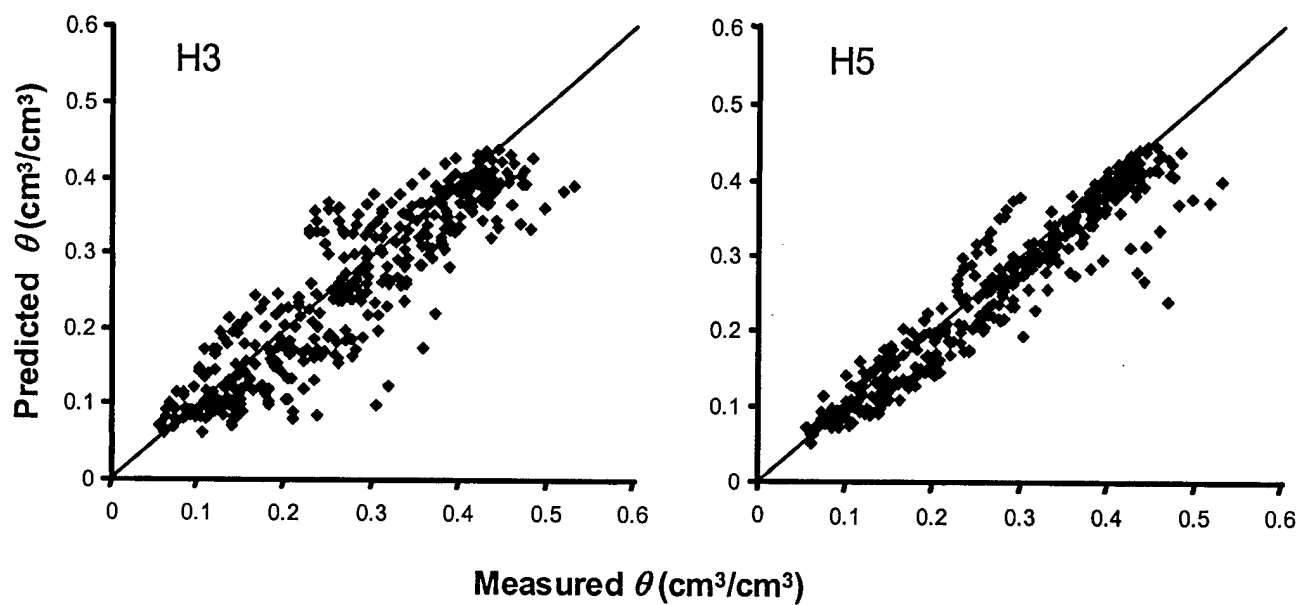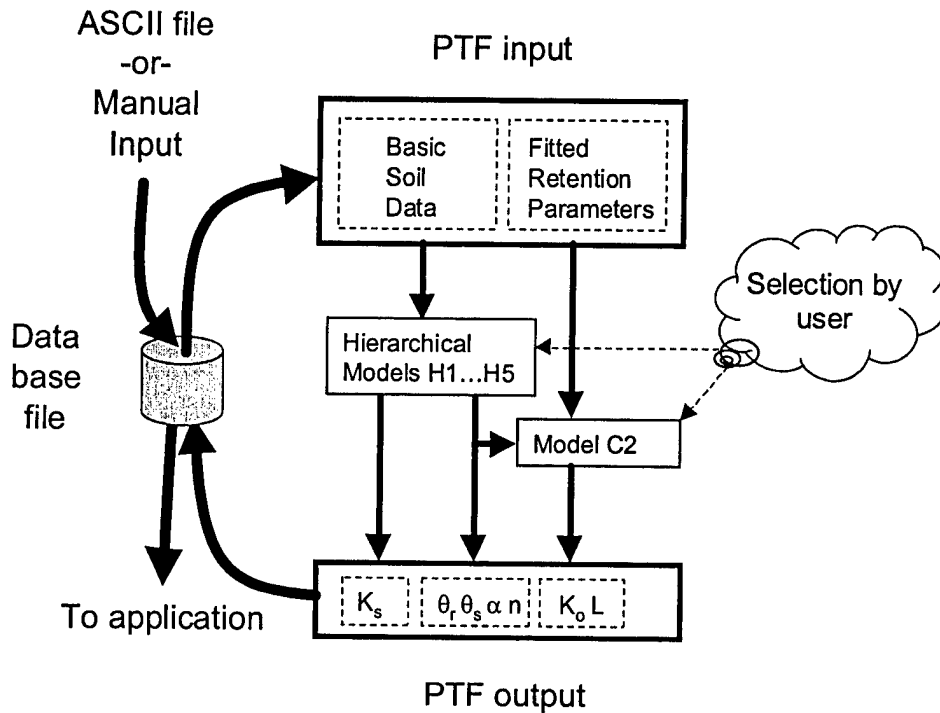


**Figure 9.** Predicted versus measured water contents for model H3 and H5 for 47 loam samples with bulk densities between 1.3 and 1.4 gr/cm$^3$ (cf. Fig 8).

The ability of the hydraulic functions to match the hydraulic data is another important factor. Figures 4 and 5 demonstrated that fits of Eq. (2) described water retention data well. For hydraulic conductivity however, we see that the direct fit of Eq. (3) has large $RMSE_u$ near saturation and a predominantly negative $ME_u$ (Fig. 6 and 7). This indicates that Eq. (3) is not suitable for simultaneously fitting the wet and the dry part of the unsaturated hydraulic conductivity curve. All models based on fitted $K_o$ and $L$ will perform poorly near saturation. *Schaap and Leij* [2000] found that the fitted $K_o$ was often about one order of magnitude lower than the measured $K_s$ value, thus causing a discontinuity in hydraulic conductivity if Eq. (3) near saturation. They interpreted this difference in terms of macropores that predominantly influenced $K_s$ and properties of the soil matrix that determined $K_o$. An improved version of Eq. (3) may need to consider the effects of macropores. Unfortunately, such an effort is hampered by a lack of measured hydraulic conductivities near saturation (cf. Fig. 6).

**Description of Rosetta**

The program is named somewhat whimsically after the Rosetta Stone that allowed translation of ancient Egyptian hieroglyphs into Old Greek, the computer program Rosetta has a somewhat more mundane task of enabling the user to "translate" basic soil data into soil hydraulic properties. Rosetta allows user-friendly access to models H1…H5 for water retention and saturated hydraulic conductivity and models C2-Fit and C2-H1…C2-H5 for unsaturated hydraulic conductivity. This section can only describe the most important features of Rosetta. More information about various aspects of the program and file specifications may be obtained through the help system which can be accessed from anywhere within the program. Rosetta is primarily a Windows 98® application but should work with upgraded versions of Windows 95® and NT® as well. Data and predictions are stored in a Microsoft ACCESS 97® database file; The Microsoft ACCESS software, however, is not needed to run Rosetta. Command-line versions of Rosetta (without data base support) are available for the MS-DOS® and the LINUX operating systems. Rosetta is freely available at: http:\\www.ussl.ars.usda.gov\models\rosetta\rosetta.htm.

The flow of data inside Rosetta is illustrated in Figure 10. Input data can be entered manually screens or be input through a formatted ASCII file. Input and output data are stored in various tables in the same database file and accessed by the program as needed. Basic soil data (sand, silt, and clay percentages, bulk density and the water contents at 330 and 15,000 cm) are used by the hierarchical models (H1…H5) to predict water retention parameters and $K_s$. Model C2 uses fitted retention parameters to predict $K_o$ and $L$ (i.e. C2-Fit) but is also able to use predicted retention as input as illustrated by the model combinations C2-H1 through C2-H5 in this study. Predicted hydraulic parameters are displayed on the screen and stored in the database file along with the input data.

16

ASCII file
-or-
Manual
Input

PTF input

Basic Soil Data

Fitted Retention Parameters

Data base file

Hierarchical Models H1...H5

Selection by user

Model C2

To application

$K_s$

$\theta_r\,\theta_s\,\alpha\,n$

$K_o\,L$

PTF output

**Figure 10.** Schematic overview of the structure of Rosetta.

After starting Rosetta, a user will typically open an existing database file or create a new one. Using the menu bar, the user will then probably select the hierarchical models to predict all seven hydraulic parameters (i.e. $\theta_r$, $\theta_s$, $\alpha$, $n$, $K_s$, $K_o$ and $L$) or model C2-Fit to predict $K_o$ and $L$ using fitted retention parameters according to *Schaap and Leij* [2000]. Both options will open different screens, but for reasons of brevity, we will deal here only with the hierarchical models.

The screen for the hierarchical models displays data ("records") for one soil sample at the time (Figure 11). At the bottom, the user can select one out of six options, five of which correspond to models H1...H5. The "Best possible model" is selected by Rosetta based on the available input data. The box on the left marked "Input Data" allows manual entry or revision of input data. The fields near "Code" (top of the box) identify particular records. The other fields allow input of textural data, bulk density and one or two water contents. Depending on the model that was selected, the appropriate fields will light up for data entry. The user can only select the textural class if model H1 is selected (bottom of Fig. 11). For the other models, this box will contain an automatic (USDA) texture classification based on the sand, silt, and clay percentages. Records can be added and deleted with the "+" and "-" symbols; ongoing modifications of records can be abandoned by clicking on the "🖑" symbol. Besides manual entry, data can also be entered by means of formatted ASCII files through the "File" menu bar. We suggest this mechanism when more than a few soil samples need to be entered. The four "arrow" signs on the left side of the menu bar allow navigation through the contents of the database.

17

**Figure 11.** Screen view of Rosetta's hierarchical models, see text for explanation

Once the data are entered, predictions can be made by clicking the mouse on the single or double exclamation marks in the toolbar. The single exclamation mark will generate predictions for the currently visible record, while the double exclamation mark will generate predictions for the entire database. The latter option may take some time, depending on the number of records in the database and the processor speed of the computer. The user will be notified of problems with the input data (e.g. clay+silt+sand > 100%) when the single exclamation mark is used. Invalid input data result in "-9.9" in the output fields.

Predicted hydraulic properties appear in the box marked with "Output Data". The top of this box identifies the model used for the prediction by its input data (i.e., sand, silt, clay, bulk density, and $\theta$ at 330 and 15,000 cm in Fig 11). The predicted hydraulic parameters appear along with their uncertainties as standard deviations. These standard deviations are based on the variability in predictions among the 60 or 100 neural networks inside each model (see section 2.3). The parameters $\alpha$, n, $K_s$ and $K_o$ are predicted as $\log_{10}$ values and the uncertainties apply to these figures. The standard deviations should not be interpreted as "field variability", rather they are model uncertainties that depend on the calibration data set [cf. *Schaap and Leij*, 1998].

The predicted hydraulic parameters are stored in the database and are overwritten each time a new prediction is made. Besides reading predicted properties from the screen, the user may extract input and output data from the database into a simple ASCII format. It is also possible to use the database directly in Microsoft ACCESS if so desired.

18

## Summary and Conclusions

This report first documents our efforts to establish PTFs for the unsaturated hydraulic conductivity from *fitted* retention parameters. This requires the availability of observed retention data. The major part of the report is devoted to the implementation of neural networks for the hierarchical prediction of hydraulic properties with the program Rosetta. This program contains five PTFs for prediction of water retention parameters, saturated and unsaturated hydraulic conductivity. The models were characterized in terms of their calibration data sets and the quality of their predictions. For the prediction of water retention and saturated hydraulic conductivity, it turned out that the hierarchical models performed reasonably well if many predictors were used (texture, bulk density and one or two retention points). Although the predictions were less accurate when fewer predictors were used, such models may still have more value than no predictions at all. This study showed that it was also possible to get reasonable predictions of unsaturated hydraulic conductivity using *predicted* retention parameters derived from the hierarchical models. Although the new models are better than the "traditional" Mualem-van Genuchten model ($K_o=K_s$, $L=0.5$) they are unable to deal with the transition from saturated to unsaturated hydraulic conductivity.

The models were implemented in a computer program Rosetta. This program has a user-friendly graphical interface and combines the PTFs with a simple database management structure to facilitate data management. The program is available free of charge from the GEBJ Salinity Laboratory's World-Wide-Web site.

The PTFs in Rosetta still require input data and if these data cannot be gleaned from existing databases, they must be measured in some way. Users may ask themselves whether it is preferably to devote experimental efforts to obtain input data for PTFs or to directly determine the unsaturated hydraulic properties. The answer may depend on the application. If the user is content with a moderate level of accuracy then PTFs may indeed provide hydraulic properties without further measurements. However, we showed in this study that the inclusion of one or two (measured) water contents in the input data generally leads to superior PTFs. It therefore seems that measurements cannot always be avoided if accurate predictions of hydraulic properties are required. Ultimately, however, the accuracy of hydraulic properties cannot be a goal in itself. Rather, the context in which they are used will define the required accuracy.

Quantifying the unsaturated hydraulic properties with direct and indirect methods will benefit greatly if an adequate database of (potential) input data and output data is available. In the past we have developed the UNSODA database [*Leij et al.*, 1996]. As part of the project, we have converted this database into a Windows version [cf. *Nemes et al.*, 2001]. Just as Rosetta, the database can also be acquired from our website http:\\www.ussl.ars.usda.gov.

Finally, the PTF concept may also be extended to predict parameters for chemical transport from more readily available data. Measurement of transport parameters is hampered by similar constraints as those for hydraulic properties. We have lately developed a neural network to predict transport parameters (dispersion coefficient, solute retardation factor, and nonequilibrium parameters) based on a data set for Portuguese clay soils (reference 5).

A remaining and important challenge is the development of PTFs that account for the calibration and application scale. Thus far, our neural network models were based on "point" samples. However, many potential applications involve input and output data over larger scales. For applications that deal with Geographic Information System and remote sensing, among others, it is desirable to develop PTFs that consider the uncertainty due to spatial variability.

# References

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

Arya, L.M., and J.F. Paris. 1981. A physico-empirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. *Soil Sci. Soc. Am. J.* 45:218-1227.

Arya, L.M., F.J. Leij, M.Th. van Genuchten, and P.J. Shouse. 1999a. Scaling parameter to predict the soil water characteristic from particle-size distribution data. *Soil Sci. Soc. Am. J.* 63:510-519.

Arya, L.M., F.J. Leij, and P.J. Shouse, M.Th. van Genuchten. 1999b. Relationship between the hydraulic conductivity function and the particle-size distribution. *Soil Sci. Soc. Am. J.* 63:1063-1070.

Bouma, J. and J.A.J. van Lanen. 1987. Transfer functions and threshold values: From soil characteristics to land qualities. p 106-110. *In* K.J. Beek et al. (ed.). Quantified land evaluation. International Institute Aerospace Surv. Earth Sci. ITC publ. 6. Enschede, the Netherlands.

Burdine, N.T. 1953. Relative permeability calculations from pore size distribution data. Trans. AIME 198:71-77.

Carsel, R.F., and R.S. Parrish. 1988. Developing joint probability distributions of soil water retention characteristics. *Water Resour. Res.* 24:755-769.

Efron, B., and R.J. Tibshirani. 1993. An Introduction to the bootstrap. Monographs on statistics and applied probability. Chapman and Hall, New York, NY.

Gonçalves, M. C., F. J. Leij, and M. G. Schaap. 2001. Pedotransfer Functions for Solute Transport Parameters. *European Journal of Soil Science* (in press).

Haverkamp, R., and J.Y. Parlange. 1986. Predicting the water-retention curve from particle size distribution: 1. Sandy soils without organic matter. *Soil Sci.* 142:325-339.

Hoffmann-Riem, M. Th. van Genuchten, and H. Flühler. 1999. A general model of the hydraulic conductivity of unsaturated soils. *In*: van Genuchten, M.Th., F.J. Leij, and L. Wu (eds), Proc. Int. Workshop, Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media, pp 31-42, University of California, Riverside, CA.

Klute, A. 1986. ed. Methods of soil analysis, Part 1: Physical and mineralogical methods, second Edition. *Monogr. 9.* ASA and SSSA, Madison, WI.

Kosugi, K. 1999. General model for unsaturated hydraulic conductivity for soils with lognormal pore-size distribution. Soil Sci. Soc. Am. J. 63:270-277.

Leij, F. J., W. J. Alves, M. Th. van Genuchten and J. R. Williams, 1996. Unsaturated Soil Hydraulic Database, UNSODA 1.0 User's Manual. Report EPA/600/R-96/095, U.S. Environmental Protection Agency, Ada, Oklahoma. 103 pp.

Leij, F.J., W.R. Russell, and S.M. Lesch. 1997. Closed-form expressions for water retention and conductivity data. *Ground Water* 35:848-858.

Leij, F.J. and M. Th. van Genuchten, 1999. Characterization and measurement of the hydraulic properties of unsaturated porous media. *In*: van Genuchten, M.Th., F.J. Leij, and L. Wu (eds), Proc. Int. Workshop, Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media, pp 31-42, University of California, Riverside, CA.

Messing, I. 1989. Estimation of the saturated hydraulic conductivity in clay soils from soil moisture retention data. *Soil Sci. Soc. Am. J.* 53:665-668.

Minasny, B., A.B. McBradney, and K. Bristow. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma.* 93:225-253.

Mualem, Y. 1976. A new model predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* 12:513-522.

Nelder, J.A. and R. Mead. 1965. A simplex method for function minimization. *Computer J.* 7:308-313.

Nemes, A., M.G. Schaap, F.J. Leij, and J. H. M. Wösten. 2001. Structure and use of the UNSODA hydraulic databse, V2.0. *Journal of Hydrology.* (submitted)

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

Pachepsky, Ya.A., D. Timlin, and L.R. Ahuja. 1999. Estimating saturated hydraulic conductivity using water retention data and neural networks. *Soil Sci.* 164:552-560.

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

Rawls, W.J. and D.L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling. p 293-299. *In* Jones, E.B. and T.J. Ward (ed.) Watershed management in the eighties. Proc. Irrig. Drain. div., ASCE, Denver, CO. April 30 - May 1. 1985. 34:3293-3302.

Rawls, W.J., T.J. Gish, and D.L. Brakensiek. 1991. Estimating soil water retention from soil physical properties and characteristics. In: Stewart, B.A. *Advances in Soil Science*, Springer-Verlag, New York.

Richards, L. A. 1931. Capillary conduction of liquids through porous mediums. *Phys.* 1:318-333.

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

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten 1998. Neural network analysis for hierarchical prediction of soil water retention and saturated hydraulic conductivity. *Soil Sci. Soc. Am. J.* 62:847-855.

Schaap, M.G., and F.J. Leij, 1998. Database related accuracy and uncertainty of pedotransfer functions, *Soil Science* 163:765-779.

Schaap, M.G., and F.J. Leij. 2000. Improved Prediction of Unsaturated Hydraulic Conductivity with the Mualem-van Genuchten Model, published by *Soil Sci. Soc. Am. J.* 64:843-851.

Soil Survey Staff, 1995. Soil survey laboratory information manual. Soil Surv. Invest. Rep. no 45. Natl. Soil Surv. Center, Lincoln, NE.

Schaap, M. G., F. J. Leij, and M. Th. Van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol..*

Tyler, S.W., and S.W. Wheatcraft. 1989. Application of fractal mathematics to soil water retention estimation. *Soil Sci. Soc. Am. J.* 53:987-996.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Am. J.* 44:892-898.

van Genuchten, M.Th., and F.J. Leij. 1992. On estimating the hydraulic properties of unsaturated soils. p 1-14. *In* M. Th. van Genuchten, M.Th., F.J. Leij, and L.J. Lund (ed.). Indirect

methods for estimating the hydraulic properties of unsaturated soils. Proc. Int Worksh. Riverside, CA. Oct 11-13. 1989. University of California, Riverside, CA.

Vereecken, H., J. Maes, J. Feyen, and P. Darius. 1989. Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. Soil Sci. 148:389-403.

Williams, R.D., L.R. Ahuja, and J.W. Naney. 1992. Comparison of methods to estimate soil water characteristics from limited texture, bulk density, and limited data. *Soil Sci.* 153:172-184.

Wösten J.H.M., P.A. Finke and M.J.W. Jansen. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* 66:227-237.

Yates, S.R., M.Th. van Genuchten, A.W. Warrick, and F.J. Leij. 1992. Analysis of measured, predicted, and estimated hydraulic conductivity using the RETC computer program, *Soil Sci. Am. J.* 56:347-354.