Weighted Recalibration of the Rosetta Pedotransfer Model with Improved Estimates of Hydraulic Parameter Distributions and Summary Statistics (Rosetta3)

Yonggen Zhang and Marcel G. Schaap

Department of Soil, Water and Environmental Science, the University of Arizona, Tucson, AZ
85721, USA

Corresponding author:

Marcel G. Schaap (mschaap@cals.arizona.edu)

Telephone number: +1-520-626-4532

Fax Number: +1-520-621-1647

Abstract

Pedotransfer functions (PTFs) have been widely used to predict soil hydraulic parameters in favor of expensive laboratory or field measurements. Rosetta (Schaap et al., 2001, denoted as Rosetta1) is one of many PTFs and is based on artificial neural network (ANN) analysis coupled with the bootstrap re-sampling method which allows the estimation of van Genuchten water retention parameters (van Genuchten, 1980, abbreviated here as VG), saturated hydraulic conductivity (K_s), and their uncertainties. In this study, we present an improved set of hierarchical pedotransfer functions (Rosetta3) that unify the water retention and K_s submodels into one. Parameter uncertainty of the fit of the VG curve to the original retention data is used in the ANN calibration procedure to reduce bias of parameters predicted by the new PTF. One thousand bootstrap replicas were used to calibrate the new models compared to 60 or 100 in Rosetta1, thus allowing the uni-variate and bi-variate probability distributions of predicted parameters to be quantified in greater detail. We determined the optimal weights for VG parameters and K_s , the optimal number of hidden nodes in ANN, and the number of bootstrap replicas required for statistically stable estimates. Results show that matric potential-dependent bias was reduced significantly while root mean square error (*RMSE*) for water content were reduced modestly; *RMSE* for K_s was increased by 0.9% (H3w) to 3.3% (H5w) in the new models on log scale of K_s compared with the Rosetta1 model. It was found that estimated distributions of parameters were mildly non-Gaussian and could instead be described rather well with heavy-tailed α -stable distributions. On the other hand, arithmetic means had only a small estimation bias for most textures when compared with the mean-like “shift” parameter of the α -stable distributions. Arithmetic means and (co-)variances are therefore still recommended as summary statistics of the estimated distributions. However, it may be necessary to parameterize the distributions in different ways if the new estimates are used in stochastic analyses of vadose zone flow and transport. Rosetta1 and Rosetta3 were implemented in the python programming language, and the source code as well as additional

documentation is available at: <http://www.cals.arizona.edu/research/rosettav3.html>.

Keywords: pedotransfer function; vadose zone; hydraulic parameter; neural network; soil; water content

1 Introduction

The quantification of movement of water and dissolved solutes between the soil surface and the groundwater table requires accurate knowledge of hydraulic properties (Assouline and Or, 2013; Rawls et al., 1991; Šimůnek et al., 2008; Vrugt et al., 2008; Zhang et al., 2016). Direct measurement of these properties continues to be expensive, difficult, and time-consuming partially owing to the strong non-linearity of the water retention and unsaturated hydraulic conductivity functions and partially because hydraulic properties are subject to substantial spatial variability (Neuman, 2005, 1982; Sudicky, 1986). Over the past several decades, pedotransfer functions (PTFs) have therefore served as viable inexpensive and rapid ways to determine soil hydraulic properties (Dai et al., 2013; Li et al., 2007; McBratney et al., 2002; Minasny and McBratney, 2002; Pachepsky et al., 2006, 1996; Pachepsky and Rawls, 1999; Pringle et al., 2007; Rawls et al., 1991; Rawls and Brakensiek, 1985; Romano and Palladino, 2002; Schaap et al., 2004, 2001; Schaap and Bouten, 1996; Schaap and Leij, 2000, 1998a, 1998b; Twarakavi et al., 2009; Wösten et al., 1999). Useful reviews regarding PTFs can be found in Wösten et al. (2001), Pachepsky and Rawls (2004), and Vereecken et al. (2010), among others.

Rosetta (Schaap et al., 2001), is one of the many PTFs that can be found in the literature. While Rosetta (denoted as Rosetta1) is not necessarily better than other PTFs when tested on large datasets (see, for example, Schaap et al., 2004 and Vereecken et al., 2010), it is somewhat unique in the field of vadose zone hydrology in that it uses a hierarchical approach (described below) to suit particular cases of input data-availability to estimate water retention parameters in van Genuchten's (1980) equation (VG), saturated hydraulic conductivities (K_s) and unsaturated hydraulic conductivity (Schaap et al.,

2001). Another feature is that Rosetta can provide univariate uncertainties of its estimates. Due to the use of artificial neural networks (ANNs) as well as the bootstrap method (Efron and Tibshirani, 1993), the exact mathematical structure of Rosetta was never explicitly published (Rosetta consists of hundreds of matrices, each with tens of coefficients). Instead, Rosetta1 was released in 2001 as a Graphical User Interface (GUI) based Windows 98/XP application (Pachepsky and Schaap, 2004; Schaap et al., 2001). The GUI has become unsupportable, and the only current practical use of Rosetta1 is through the Hydrus1D and 3D applications (Šimůnek et al., 2012, 2008) and limited informal releases of source code by the authors. The continued interest in Rosetta1 has prompted us to consider whether the model can be improved and can be released in an alternative form.

The models available in Rosetta1 have several deficiencies. Firstly, as was already demonstrated in Schaap et al. (2001) that there was a substantial pressure head-dependent bias in estimated water contents. Depending on the model, Rosetta1 typically underestimates water contents by 0.02 to 0.05 cm^3/cm^3 at pressure heads larger than 100 cm (see Figure 3, Schaap et al., 2001 or Figure 6a in the present work). This problem finds its roots in the fact that the objective function for the calibration of the ANNs was formulated in terms of dimensionless hydraulic parameters. Evaluation metrics, however, were formulated in terms of water contents and because water contents are a non-linear function of the VG parameters, an optimum in terms of hydraulic parameters does not necessarily mean that the best metrics in terms of water contents are achieved as well. As shown in Figure 3 in Schaap et al. (2004), it is possible to correct for the pressure head-dependent bias by establishing a linear transformation for each estimated VG parameter (this model is denoted as Rosetta2). However, this correction is somewhat ad-hoc and cumbersome and we will demonstrate an alternative method in this study that does not require a posterior “correction” to the PTF estimates.

Secondly, uncertainty in the *fitted* VG parameters that were used to calibrate models in Rosetta1 was ignored in Schaap et al. (2001). This implies that VG parameters fitted to noisy incomplete retention characteristics (e.g. those where low water contents are missing due to experimental

constraints) carried the same weight as VG parameters fitted to “perfect” smooth and complete characteristics. In the former case, some or all fitted parameters carry a large uncertainty, while in the latter case the parameter uncertainty is much smaller. A hypothesis in this study is that the performance of the PTFs in Rosetta1 can be improved by accounting for the uncertainty of fitted VG parameters while calibrating the ANN models. However, we will also demonstrate that the parameter uncertainties cannot be taken simply at face-value but need to be constrained to arrive at a model that is indeed better than previous work.

Thirdly, while the PTFs in Rosetta1 can provide uncertainties for estimated parameters, they can provide only *univariate* uncertainties owing to the fact that previously separate models were developed for the van Genuchten (1980) water retention parameters and K_s . A full (five-parameter) covariance matrix for water retention parameters and K_s cannot be generated. Such a matrix is useful for stochastic studies where the correlation among hydraulic parameters is relevant. In addition, computational limitations faced by Schaap et al. (2001) limited the bootstrapped models to 60 and 100 for water retention and K_s models, respectively. A larger number of bootstrap replicas would be useful for evaluating whether Rosetta1’s estimates have Gaussian distributions (as assumed by Schaap et al., 2001) or enhance the heavy-tailed nature of textural distributions in some sediments (such as found by Guadagnini et al., 2013).

Finally, as mentioned above, the current published Rosetta1 software has been rendered nearly unusable because of the challenges posed by supporting a legacy operating system-specific source code. It is desirable that both the Rosetta1 and Rosetta3 models remain available in the form of open-source code in a major programming language with a likelihood of long-term support.

The objective of this study is therefore to recalibrate the Rosetta1 model in such a way that one model can predict van Genuchten water retention parameters and K_s and produce a full (five-parameter) covariance matrix. In order to provide continuity with the past and allow for a good benchmark, we will recalibrate the new models on the same dataset as Schaap et al. (2001). However,

the proposed methodologies are sufficiently general that they can be used for PTFs that use different hydraulic models or calibration datasets. In order to maximize model performance we will also investigate whether ANN topology used in Rosetta1 must be changed. In addition, we will also determine the number of bootstrap realizations needed to arrive at stable model results.

104

105 **2 Materials and methods**

106 **2.1 Dataset and soil hydraulic parameters**

The dataset used in this study is the same as used by Schaap et al. (2001), which contains 2134 soil samples for water retention with a total of 20,574 retention points, and a subset of 1306 samples with valid data for K_s . Soil water retention is described with van Genuchten's (1980) equation:

$$110 \quad \theta(h) = \begin{cases} \theta_r + \frac{\theta_s - \theta_r}{\left[1 + |\alpha h|^n\right]^m} & h \leq 0 \\ \theta_s & h > 0 \end{cases} \quad [1]$$

where $\theta(h)$ is volumetric water content ($\text{cm}^3\text{cm}^{-3}$) at pressure head h (cm, in this study taken positive for unsaturated conditions); θ_r and θ_s are residual and saturated water contents, respectively ($\text{cm}^3\text{cm}^{-3}$); α (>0 , in cm^{-1}) and n (>1) are curve-shape parameters, and $m=1-1/n$. The parameters θ_r , θ_s , α and n are designated as “VG parameters”.

115 **2.2 Fitting of hydraulic parameters and associated weights**

A non-linear parameter optimization procedure (the NL2SOL algorithm in “nls” function from the statistical package R, version 3.0.2, Ihaka and Gentleman, 1996; Venables and Smith, 2003) is used to optimize VG parameters for each soil sample by minimizing:

$$119 \quad \chi^2(\mathbf{p}) = \sum_{i=1}^{N_w} [\theta_i - \theta'_i(h_i)]^2 \quad [2]$$

where θ_i , $\theta'_i(h_i)$ are measured and estimated water content of the i^{th} water retention data, respectively; N_w is the number of measured water retention data for each soil sample; $\mathbf{p}$ is the VG parameter vector

122 $(\theta_r, \theta_s, \alpha, n)$, which is used to estimate $\theta'_i(h_i)$ using corresponding h_i .

123 The optimal VG parameters are determined by iterations using the above mentioned optimization
124 procedure from a given initial guess of VG parameters. Once determined, a *formal* covariance matrix
125 (Press et al., 1992, p695) associated with VG parameters can be calculated from:

$$126 \quad \mathbf{Cov}_{VG} = [\mathbf{J}^T \mathbf{J}]^{-1} \quad [3]$$

127 where $\mathbf{J}$ is a Jacobian matrix with dimension of $N_w \times 4$, whose components are constructed from the
128 derivatives of state variables with respect to each of the four VG parameters at N_w observations.

129 The corresponding weighting matrix of VG parameters is taken as the inverse of the diagonal
130 matrix of the covariance matrix $\mathbf{Cov}_{VG}$:

$$131 \quad \mathbf{W}_{VG} = [\text{diag}(\mathbf{Cov}_{VG})]^{-1} \quad [4]$$

132 Fitted parameters with large variances are therefore weighted less than ones with small variances.

133 2.3 Constraints on parameter weights

134 The covariance matrix defined by Eq. [3] provides an inaccurate estimate of the *true* variance or
135 covariance pairs because of 1) measurement errors and whether these errors are, in fact, Gaussian, 2)
136 model error (Schweppe, 1973, p. 120): the VG equation is an empirical formula may not always
137 describe retention well, and 3) because the VG equation is a non-linear function of pressure head.
138 Indeed, Eq. [4] leads to weights that have an unreasonably large variation across 10 orders of
139 magnitude. Figure 1 depicts the sorted *cumulative* contribution of weights for the four VG parameters
140 (scaled to 100%). In the case of θ_r , the sum of weights of 1451 samples contribute to less than 1% of
141 the total cumulative weight. By contrast, 81 samples with the largest weights were responsible for 80%
142 of the total cumulative weight. The weights for θ_s , α , and n had similar issues (Figure 1b, 1c, and 1d).
143 The above implies that most of the weight is carried by a limited number of samples. In addition, the
144 fraction of samples that carry 90% or 99% of weights tend to have a statistically different textural

145 distributions (as determined by a Kruskal-Wallis test, not shown) compared against samples with low
 146 weights.

147 Taking the weights according to the definition in Eq. [4] will therefore lead to a PTF that is biased
 148 in texture and that is not representative for the entire dataset. Our solution to this problem is to *limit* the
 149 weights relative to the median weights for each parameter according to:

$$150 \quad w'_{ij} = \begin{cases} w_{j,median} \times \Omega_U & \text{if } w_{ij} > w_{j,median} \times \Omega_U \\ w_{j,median} \times \Omega_L & \text{if } w_{ij} < w_{j,median} \times \Omega_L \end{cases} \quad [5]$$

151 where w'_{ij} denotes the adjusted weight of j^{th} VG parameters for sample i , while w_{ij} is the corresponding
 152 original weight determined from Eq. [4]; $w(x)_{j,median}$ is the median value of the weight for the j^{th} VG
 153 parameters. The empirical parameters Ω_U and Ω_L are the upper and lower cut off values of the
 154 weights; their values will be determined and discussed in Section 3. In addition, the weights must be
 155 normalized using:

$$156 \quad w''_{ij} = \frac{w'_{ij}}{w_{j,median}} \quad [6]$$

157 which is subsequently used in Section 2.4.

158 **2.4 Model development**

159 We developed five new models using the same hierarchical approach as Schaap et al. (2001) did.
 160 The first model (H1w) is a simple lookup table for each USDA soil textural class, while the second
 161 model (H2w) to the fifth model (H5w) are ANNs combined with bootstrap re-sampling, which can
 162 provide hydraulic parameters that vary continuously with soil texture.

163 H1w provides weighted parameter averages and weighted standard deviations of soil hydraulic
 164 parameters for each USDA soil textural class according to

$$165 \quad \bar{p}_{H1,j} = \frac{\sum_{i=1}^{N_c} w''_{ij} p_{ij}}{\sum_{i=1}^{N_c} w''_{ij}} \quad [7]$$

$$SD_{H1,j} = \sqrt{\frac{\sum_{i=1}^{N_c} w_{ij}''}{\left(\sum_{i=1}^{N_c} w_{ij}''\right)^2 - \sum_{i=1}^{N_c} w_{ij}''^2} \sum_{i=1}^{N_c} w_{ij}'' (p_{ij} - \bar{p}_{H1,j})^2} \quad [8]$$

where $\bar{p}_{H1,j}$ and $SD_{H1,j}$ are weighted mean and weighted standard deviation in H1w model for the j^{th} hydraulic parameters of each textural class; N_c is the number of available soil samples for which VG and K_s in each USDA soil textural class are calculated.

The second model (H2w) uses soil texture (sand, silt, and clay percentages) as input data and provides hydraulic parameters that vary continuously with soil texture. Bulk density is included as an additional input for the third model (H3w). The fourth model (H4w) also uses the water content at 330 cm pressure head, while the last model (H5w) includes the water content at 15,000 cm pressure in addition to the input for H4w. The choice of pressure heads in models H4w and H5w was motivated by their availability in the USDA National Cooperative Soil Survey Characterization (NCSS) database (National Cooperative Soil Survey, 2016). In the present work, we maintained the requirement of sand, silt and clay percentages, even though one of these variables is redundant (because they sum to 100%). Preliminary analyses revealed that model performance was not improved when only two variables were used while it would also be more difficult to perform a sanity check on the data (i.e. that sand, silt and clay indeed sum to 100%).

We used the function “feedforwardnet” neural network in Matlab R2015a to develop the ANN models and refer to the documentation for details (Mathworks, 2015). The ANN topology was similar to that discussed in Schaap and Bouten (1996) and Schaap and Leij (1998b) and consisted of a simple feedforward network with one hidden layer and six hidden nodes. The Rosetta1 model used sigmoidal transfer functions in the hidden and output layers used “tansig” and “purelin” for hidden and output nodes, respectively. In this study, we found that a combination of “tansig” function for the hidden nodes and “logsig” function for the output nodes provided somewhat better results (results not shown).

188 Custom MATLAB code was written to bootstrap the H2w through H5w models; model H1w was
189 not bootstrapped because it had a limited number of samples in some classes. Briefly, the bootstrap
190 method (Efron and Tibshirani, 1993) generates multiple random subsets by sampling from the original
191 dataset leading to pairs of complimentary calibration-validation (or “replica”) datasets that can be used
192 for calibration of the ANNs as well as validation of the individual ANNs. Bootstrap randomly selects
193 samples with replacement, which results in $1 - [(N-1)/N]^N$ of the samples (around 63%) to be present
194 in any given replica dataset used for calibration. The remaining 37% of the dataset can therefore be
195 used to validate the ANN that was calibrated on the replica data (for more information see Schaap and
196 Leij, 1998a; Schaap et al., 1998). Following Hastie et al., (2001) the calibration dataset divided by
197 bootstrap is again split into training, validation, and testing with 70%, 15%, and 15% samples
198 respectively using the MATLAB “net.divideParam” function which also decides to terminate the
199 calibration of the ANN once its performance on validation data is optimal.

200

201 The bootstrap method served two purposes in our work. Firstly, it was used to establish
202 preliminary models to determine the optimal bounds on weights, the optimal weights for VG and K_s ,
203 and the optimal ANN topology. In this stage we used only 100 bootstrap replicas, mainly for
204 computational efficiency. Many model configurations were tested, and we omitted some of tedious
205 (and uninformative) results because these did not alter the main findings reported in this work.
206 Secondly, the final models (H2w through H5w) were calibrated with 1000 bootstrap replicas and
207 therefore each consist of 1000 ANN models. Each estimate with models H2w through H5w therefore
208 yields 1000 estimates of VG parameters and K_s ; model H1w just provides a mean and standard
209 deviation for each parameter and soil textural class.

210 The H2w through H5w models were calibrated using the following objective functions:

$$O_{nn}(c_{nn}) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_p} w_{ij}'' (p_{ij} - p'_{ij})^2 \quad [9]$$

where c_{nn} represents the coefficient matrices of the artificial neural network. p_{ij} and p'_{ij} are fitted / observed and estimated soil hydraulic parameters, respectively. N_s is the number of calibration samples selected from bootstrap sampling process. The number of hydraulic parameters, N_p , equals 5, including four VG parameters and K_s . w_{ij}'' represents the weights for the i^{th} calibrated sample and j^{th} hydraulic parameters determined in Eq. [6]. The weights for K_s parameters are given as a constant value for samples that have K_s measurements; they are set to zero for samples without K_s measurements.

For completeness, we note that the term *fitted* parameter indicates the optimal VG parameters determined from Eq. [2], while an *observed* quantity indicates a measured water content or K_s . The term *estimated* indicates output from the PTFs and can apply to the estimated distribution of hydraulic parameters, or summary statistics thereof.

2.5 Characterization of the distributions of the estimated parameters

For most practical purposes, the distributions of hydraulic parameters estimated by models H2w through H5w should be simplified using summary statistics such as means and standard deviations. Such summary statistics imply non-skewed Gaussian distributions and we used the Shapiro-Wilk test (Shapiro and Wilk, 1965) to determine whether the estimated distributions were indeed Gaussian. The estimated distributions were also characterized with the (Levy) α -stable model, which has the ability to describe a family of Gaussian as well as non-Gaussian distributions. An α -stable probability distribution is characterized by four parameters: stability index $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\sigma > 0$, and shift μ . Following Samoradnitsky and Taqqu (1994), and Guadagnini et al. (2013, 2014), we write its characteristic function as:

$$\ln \langle e^{i\phi Z} \rangle = i\mu\phi - \sigma^\alpha |\phi|^\alpha [1 + i\beta \text{sign}(\phi) \omega(\phi, \alpha)];$$

$$\omega(\phi, \alpha) = \begin{cases} -\tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \ln |\phi| & \text{if } \alpha = 1 \end{cases} \quad [10]$$

Here, Z is an α -stable variable, $\langle \rangle$ denotes expected value (ensemble mean), ϕ is a real-valued parameter in the characteristic function, and $\text{sign}(\phi) = 1, 0, -1$ if $\phi > 0, = 0, < 0$, respectively. The distribution becomes Gaussian for $\alpha = 2$ and $\beta = 0$, and non-Gaussian otherwise. In Section 3.4, we computed an estimate $\hat{\varepsilon} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\mu})$ of the parameter vector $\varepsilon = (\alpha, \beta, \sigma, \mu)$ by using the maximum likelihood method using the computer code STABLE developed by Nolan (2001). A detailed description of this algorithm can be found in Nolan (1997). If there is a potential for confusion between α in α -stable distributions and α in VG parameters, we will explicitly write α (in the context of α -stable distribution) or VG- α , respectively.

2.6 Evaluation criteria

The most important criteria used to quantify errors are those that evaluate the performance in measured quantities (i.e. water contents and K_s) and include root mean square error ($RMSE$) and mean error (ME) for water content and saturated hydraulic conductivity, defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\zeta'_i - \zeta_i)^2} \quad [11]$$

where ζ and ζ' indicate measured and estimated $\theta(h)$ or $\log_{10} K_s$; N is the number of measurement for water content and saturated hydraulic conductivity in the entire dataset; N equals 20,574 for water content, and 1306 for $\log_{10} K_s$. The $RMSE$ metric was also used to determine whether the α -stable distribution was a better description of the estimated distributions than a Gaussian curve (see Section 3.4 for details).

The mean error (ME) is used to quantify systematic errors between the measurements and

252 estimations as:

$$253 \quad ME = \frac{1}{N} \sum_{i=1}^N (\zeta'_i - \zeta_i) \quad [12]$$

254 Water content and saturated hydraulic conductivity are designated by the subscripts w and K_s
255 respectively for $RMSE$ and ME . ME_w values are negative (positive) when the PTF underestimates
256 (overestimates) water contents. To investigate how the ME_w values vary with pressure heads, ME_w
257 values for nine pressure classes between 0, 3.2, 10, 32, 100, 320, 1000, 3200, 10,000, and 32,000 cm
258 were also computed (see Schaap et al., 2001). This was done for the entire dataset, as well as for four
259 groups of USDA textural classes (Schaap and Leij, 2000): sands (sand, loamy sand, sandy loam, and
260 sandy clay loam), loams (loam and clay loam), silts (silt and silty loam), and clays (clay, silty clay, silty
261 clay loam, and sandy clay).

262 The correspondence between estimated and fitted or observed parameters j was calculated with the
263 weighted coefficient of determination (R^2):

$$264 \quad R^2 = 1 - \frac{\sum_{i=1}^N w_{ij}'' (p_{ij} - p'_{ij})^2}{\sum_{i=1}^N w_{ij}'' (p_{ij} - \bar{p}_j)^2} \quad [13]$$

265 where $\bar{p}_j$ is the average of p_{ij} . Bounded weights (w_{ij}'') appear in this equation because p_{ij} has
266 uncertainty due to fitting. Equations [11] and [12] do not use weights because they rely on measured
267 water contents or saturated hydraulic conductivity for which the measurement errors are unknown.

268 For each ANN calibrated on a bootstrap replica we will calculate error metrics for its calibration
269 data and its complementary validation data and aggregate these over all replicas. It is also possible to
270 simply apply each ANN on the entire 2134 sample data set, such as was done in Schaap et al. (2001).
271 While the latter evaluation is biased (because the data now include calibration and validation data), we
272 maintain this method in the present work to make a direct comparison with Rosetta1 possible.

273

274 3 Results and discussion

275 3.1 Development of the weighted ANN models (H2w...H5w)

276 In order to obtain the optimal bounds for the weights of the VG parameters (Eq. [5]), we
277 temporarily set the weights for all K_s values to zero, used six hidden nodes in ANN and 100 bootstrap
278 replicas based on the preliminary analysis of Schaap et al. (2001). We defined the lower cutoff of
279 weights (Ω_L) as the reciprocal of the upper cutoff of the weights (Ω_U) as:

$$280 \quad \Omega_U = \frac{1}{\Omega_L} \quad [14]$$

281 For a value of $\Omega_U = \Omega_L = 1$, the method reverts to Rosetta1 which used no weights. Figure 2
282 demonstrates that $RMSE_w$ (left y axis values) decreases and that ME_w (right y axis values) approaches 0
283 immediately when $\Omega_U > 1$, indicating that Rosetta1 can indeed be improved upon by using weighted
284 parameters. The lowest values for $RMSE_w$ and for the absolute value of ME_w are found near $\Omega_U = 1.8$.
285 This value is also appropriate for H2w through H5w (analyses not shown). Beyond $\Omega_U = 1.8$, $RMSE_w$
286 and ME_w values deteriorate and for $\Omega_U > 5$ the $RMSE_w$ metric becomes worse than Rosetta1, indicating
287 that the weights as defined in Eq. [4] should indeed be bounded. For the remainder of this work, $\Omega_U =$
288 1.8 is used as the optimal upper cutoff of the weights, implying that $\Omega_L = 1/1.8 = 0.56$. We also
289 investigated whether Ω_U and Ω_L could be defined independently but found only minimal improvements
290 at the cost of increased complexity (details not shown).

291 Next, models that predicted VG parameters and K_s were calibrated to determine the optimal
292 weights for K_s . Only two weights were needed: a value of 0 where K_s measurements were not available
293 (which was the case for 828 out of total 2134 samples) and an empirical optimal value for the
294 remaining 1306 samples such that the errors for K_s and water retention were at a minimum. Figure 3
295 indicates that $RMSE_w$ values were nearly independent of the weight assigned to K_s : when we changed
296 the weight for K_s in increments of 0.1 between 0 and 1, and increments of 1 between 1 and 10, a nearly
297 flat response in $RMSE_w$ was found for the H3w model. Results were similar for the other models. This

figure also shows that $RMSE_{K_s}$ decreased asymptotically with increasing weight assigned to K_s . We set $w_{K_s} = 8$ for the 1306 samples for which K_s was available (and to 0 otherwise). It is noted that many alternative preliminary models were calibrated (not shown), most of which also pointed to $w_{K_s} = 8$ as an optimal choice.

ANN model complexity can be controlled by changing the number of hidden nodes. Figure 4 shows that $RMSE_w$ validation values decreased asymptotically when the number of hidden nodes in the ANN was increased from two to ten for the H3w model. A nearly flat response was found in $RMSE_{K_s}$, with a slight increase when more than six hidden nodes were used. The results in Figure 4 indicate that the number of hidden nodes should be kept at six (as also used by Schaap et al., 2001) as an acceptable compromise between the trends in $RMSE_w$ and $RMSE_{K_s}$. We note that the results in Figure 4 pertain to $RMSE$ for validation results only because the use of calibration data can lead to unreliable results due to overfitting on calibration results.

With the ANN models defined in their final form in terms of the optimal weights and topology, Figure 5 provides the evolution of the progressive mean $RMSE_w$, $RMSE_{K_s}$, ME_w , and ME_{K_s} values for a total of 1000 bootstrap replicas (model H3w). It is seen that the error values stabilize after a few hundred bootstrap replicas. With this diagram it becomes clear that the usage of 60 (VG parameters) or 100 (K_s) in the model of Schaap et al. (2001) was probably too small. To be on the safe side, we defined each of the new models to consist of the complete set of 1000 ANNs. This number will ensure stable summary statistics and keep estimates in the order of a millisecond per sample on modern computer hardware. In addition, 1000 replicas will make Rosetta3 model useful for stochastic modelling of vadose zone and subsurface hydrology. For example, Chen and Zhang (2006), Panzeri et al. (2014), Shi et al. (2015) and Song et al. (2014) have demonstrated that 500-1000 realizations of soil hydraulic parameters can ensure reasonable ensemble statistics when using the ensemble Kalman filter.

3.2 Development of the class average model (H1w)

The bounded weights that were determined in Section 3.1 were also used to compute weighted parameter average values and weighted standard deviations of soil hydraulic parameters within each soil USDA texture class for model H1w using Eqs. [7] and [8]. Table 1 provides the H1w estimates as ratios of the H1 estimates of Rosetta1 (Schaap et al., 2001). It is noted that α , n , and K_s values were $\log_{10}$ transformed before calculating average values over each textural class, and then were anti-log transformed for the textural class average values before calculating the presented ratios. It is evident that the new method produced higher θ_t values, particularly for sandy loam, sandy clay loam, and loam classes. Minimal increased in θ_t were found for the sand samples. Substantial downward revisions are found for α for all classes, though the revisions are not as large for the sand and silt classes. Smaller and mostly downward changes are found for θ_s and n , with ratios of 0.97 to 1.00, and 0.91 to 1.00 for θ_s and n , respectively. The revisions for θ_t and α were relatively large because parameter bounds were often enforced for these parameters when the VG equation was fitted to data of truncated retention curves. This resulted in correspondingly large variances and low weights. No changes were found for K_s because this parameter had uniform-valued weights (as discussed in section 3.1) for samples with available K_s measurements. K_s estimates by model H1w are therefore identical to the previous H1 model in Rosetta1. Table 1 also provides ratios of the standard deviations of the estimates by the H1 model in Rosetta1 and the H1w model in Rosetta3 models developed in this study; α , n , and K_s values were $\log_{10}$ transformed before calculating the presented ratios. On average, the standard deviations were revised downward by about 5%, with extreme ratios of 0.8 and 1.12.

3.3 Model performance

Table 2 shows ME_w , $RMSE_w$, ME_{K_s} , and $RMSE_{K_s}$ values for corresponding calibration and validation data. The error measures are nearly identical for calibration and validation data, which indicates that our models are not overfitted. Table 3 provides the ME_w , $RMSE_w$, ME_{K_s} , and $RMSE_{K_s}$

values from Rosetta1, and the improved Rosetta3 models developed in this study. In this case all 1000 ANN models were applied to the entire dataset to maintain consistency with Table 1 in Schaap et al. (2001). ME_w values of the new models vary from 0.0051 for H1_w to 0.0017 for H5_w. These values are better than the Rosetta1 model by about one order of magnitude. $RMSE_w$ values were reduced modestly by about 5% (H1_w) to 15% (H4_w), compared to the Rosetta1 model. ME_{K_s} values were almost unchanged. The results are different for $RMSE_{K_s}$ which is *larger* in the new models by 0.9% (H3_w) to 3.3% (H5_w) compared with the Rosetta1 model. This small increase is unfortunate but is outweighed by other advantageous qualities such as allowing the uni-variate and bi-variate probability distributions of estimated parameters and substantially reduced systematic errors in retention as shown below.

Table 4 shows the calibration and validation results evaluated by weighted R^2 (Eq. 13) of Rosetta3 models. As expected, validation results are slightly inferior compared with calibration results. Table 5 shows weighted R^2 of Rosetta3 for hydraulic parameters evaluated on the entire dataset; unweighted R^2 of the Rosetta1 model are also shown. R^2 values of θ_t and $\log_{10}(\alpha)$ are improved in H1_w through H5_w model using Rosetta3 estimates compared with the Rosetta1 model. R^2 values of θ_s in H1, H2, and H5 model, and $\log_{10}(n)$ in H5 model are also improved using the new PTFs, while the match of θ_s and $\log_{10}(n)$ in the other hierarchical PTFs decreased somewhat. R^2 values of $\log_{10}(K_s)$ in the new H1_w model is almost identical compared with the Rosetta1 model, while the match of $\log_{10}(K_s)$ in the other PTFs decreased slightly. Overall, based on results shown in Tables 2 through 5 we can confidently state that the new models are better than those published in Schaap et al. (2001) and are subject to minimal overfitting issues.

Figure 6 shows how ME_w values vary with nine pressure head classes (defined in Section 2.6) in H1 through H5 in Rosetta1 models (Figures 6a - 6e) and corresponding H1_w through H5_w in Rosetta3 models (Figure 6f - 6j) for all soil samples (Figures 6a and Figures 6f respectively for Rosetta1 and

Rosetta3 models) and for soil textural groups: sands, loams, silts, and, clays (see Section 2.6 for details). The histograms provide the number of observations for each pressure head class in different soil groups. Each graph also provides the results of the direct fit (FIT) of the van Genuchten equation (Eq. [1]) to the water retention data as a benchmark value. In general, the FIT result shows no large systematic errors, indicating that water retention can be described adequately with Eq. [1]. By contrast, considerable ME_w were present for the Rosetta1 model (between -0.05 and 0.03 cm³cm⁻³). Substantially reduced systematic errors were found for the new models for pressure heads higher than 32 cm but rather large ME_w remained between saturation and 32 cm for all textural groups which are possibly due to the scarcity of observations at these pressures as indicated by the histograms. Combined with the modest improvement in $RMSE_w$, these results indicate that model calibration with bounded weights indeed leads to better estimates with less pressure head dependent bias.

3.4 Characterization of the distribution of estimates

For each set of input variables, the H2w through H5w models produce a distribution of 1000 estimates for each of the five hydraulic parameters (VG parameters and K_s). Evidence that these distributions are non-Gaussian is shown for model H2w in Figure 7b - 7f, which depicts the results of a Shapiro-Wilk test (Shapiro and Wilk, 1965) for every whole percentage point in the soil textural triangle (0..100% sand and 0..100% clay). For reference, Figure 7a shows the soil texture distribution of the original data. Textures with colored contours are the only regions where the *null* hypothesis (Gaussian distribution) could not be rejected (at p -values > 0.05). It is therefore likely that the white regions hold non-Gaussian distributions.

We elaborate the non-Gaussian nature of the estimates in Figures 8 and 9, the diagonals of which show the histograms of estimates for a clay loam sample (40% sand, 40% clay) and an extreme clay sample (5% sand, 90% clay), respectively. The first sample is from a well-populated area (Figure 7a) with likely Gaussian distribution of estimates (coloured in Figure 7c - 7e), while the second one is likely non-Gaussian. Also shown are the fits of Gaussian distributions (red lines) and α -stable

395 distribution (green lines), as well as the corresponding $RMSE$ (G for Gaussian and α for α -stable). The
396 estimated univariate distributions for the clay loam are indeed Gaussian, or nearly so, as indicated by
397 the overlapping green and red lines. By contrast, the estimated univariate distributions for the clay
398 sample are distinctly non-Gaussian with $RMSE_{\alpha} < RMSE_G$ indicating that the α -stable distribution
399 provides a better description of the histogram than does the Gaussian curve. These figures further
400 indicate that the estimated distributions are mostly symmetric and not particularly skewed. The off-
401 diagonal diagrams show the bi-variate distributions of the estimates; to show the scatter among
402 estimates, the ranges of x and y axis values of the lower triangle in Figure 9 are different from those in
403 Figure 8. Pearson correlation was low (< 0.45 , not shown) indicating a low co-variance for these
404 samples. The lower triangle of Figure 8 and 9 also show arithmetic means, medians and values of μ as
405 colored points. These values almost overlap for the clay loam (Figure 8), but show larger differences in
406 Figure 9 (clay). It is interesting to note that the median of the distributions is always very close to the
407 value of μ .

408 The non-Gaussian nature of the distributions is mostly caused by stability index $\alpha < 2$ (Eq. [10])
409 for all variables (Figure 10), which reaches particularly low values in the extreme silt and clay where
410 the calibration database has relatively few samples (see Figure 7a for details). However, the fitting of
411 α -stable distributions is computationally intensive and somewhat error-prone in practice because of
412 convergence problems. Even though α -stable distributions appear to be a more consistent way to
413 describe estimated distributions, most applications are interested in single-valued estimates of
414 hydraulic parameters rather than raw or parameterized distributions. In order to quantify the effect of
415 bias of estimated arithmetic means (which is caused by a slight skewness of the distributions), it is
416 therefore worthwhile to visualize by how much arithmetic means deviate from the quantity μ (μ is
417 equal to the arithmetic mean value if $\alpha = 2$ and $\beta = 0$) in Eq. [10]. Figure 11 provides the normalized
418 bias for each estimated parameters, which is calculated as $(\text{mean} - \mu)/\text{mean}$. A value of 0 indicates no

419 bias. The bounds in the legend, [-0.2, 0.2], represent the minimum and maximum values that the
420 normalized bias can reach. The deviations are small (generally within the range of -0.05 to 0.05 in
421 terms of the normalized bias values) in most of the textural triangle and, similar to Figure 10, the
422 strongest deviations are again found in the extreme silt and clay corners of the textural triangle that are
423 not represented well in Rosetta's calibration database (see Figure 7a) and which are textures that do not
424 occur very often in nature (see, for example, Figure 2 in Schaap et al, 2004, which presents the
425 distribution of 47 thousand soil textures in USA).

426 Even though Figures 7 through 10 indicated that the estimated distributions were predominantly
427 non-Gaussian, in practice only small errors are made when it is assumed that the distributions are
428 Gaussian (this assumption was also made in Table 2 and 3, as well as Figure 6). The authors therefore
429 recommend the continued use of arithmetic means and (co-)variance to characterize the estimated
430 distributions; median values may also provide a useful alternative but disallow the use of co-variances.
431 If the exact shape of the distribution is important (e.g. in stochastic analyses where heavy-tailing of
432 hydraulic parameters is relevant), it is recommended to characterize the distributions in other ways, as
433 appropriate for the application.

434 **3.5 Software**

435 A unique aspect of ANN-based PTFs is that their internal mathematics cannot be easily published,
436 and this is especially the case when the models are combined with the bootstrap method. The new
437 models, as well as the models of Schaap et al. (2001) have been (re-) implemented in the Python
438 programming language which ensures computing platform independence, great versatility and the
439 convenience of the code that can be integrated in existing software or even web-based applications.
440 The use of the new models consist of two software calls. The model is first instantiated as:

441
$$model = PTF_MODEL(model_number, db) \quad [15]$$

442 where PTF_MODEL is a so-called factory routine that obtains everything related to *model_number*
443 from a data base *db*. This database is distributed as a file-based SQLite database, but can also be

located on a network-based SQL database. The variable, *model_number* is a simple index that indicates H1w through H5w (index 1 through 5) of the new Rosetta3 models or the H1 through H5 (index 101 through 105) of Rosetta1 models. The object *model* holds the requested PTF model and all its functionality. Estimates can now be made with:

```
res = model.predict(data, summary_data=True) [16]
```

where *data* is a simple two-dimensional array that holds the input data that are appropriate for a particular model (e.g., sand, silt and clay percentages for H2w or H2). The object *res* holds the results as a Python dictionary. If *summary_data* is True (the default), *res* holds simple summary statistics, such as means, standard deviations and co-variances; a median is also provided. If *summary_data* is False, *res_dict* contains the raw distributions (a 5×1000 matrix) without summary statistics. It is then up to the user to further process the results.

The code has the following widely-used dependencies: Python version 2.7.x (or 3.x), Numpy (version >1.5) and SQLite3. The code is under continued development but will at least hold the models discussed in the present work; additional functionality or PTFs may be added in the future. The code and additional documentation is available under a GPLv2 open source license from: www.cals.arizona.edu/research/rosettav3.html. The code has been verified to work correctly under Linux, Windows-7 and Mac OSX. The above code is targeted mostly towards software developers who desire integration of Rosetta into existing research or commercial software. A web-based interface is also under development as an effort separate from the present work.

463

464 **4 Summary and conclusion**

The Rosetta pedotransfer function of Schaap et al. (2001) was improved by using a single artificial neural network (ANN) for all five soil hydraulic parameters (the four van Genuchten water retention parameters, θ_r , θ_s , α , and n , and the saturated hydraulic conductivity, K_s), thus allowing the uni-variate and bi-variate probability distributions of estimated parameters. The modification was accomplished

by assigning weights to each individual hydraulic parameter during the calibration. The weights for VG parameters were derived from the covariance matrix obtained from the curve-fit of VG equation to water retention data, and were empirically bounded for maximum PTF performance. The weight for the K_s parameter was set to an optimal value for samples with known K_s and to zero otherwise. Five hierarchical models were developed, similar to those of Schaap et al. (2001). The simplest model (H1w) is a lookup table that provides weighted arithmetic means and standard deviations for each USDA soil textual class. Models H2w through H5w were combined with the bootstrap method (1000 replicas), and estimates of these models consist of distributions of hydraulic parameters. We obtained the following summary results and conclusions:

- The weights of the VG parameters were derived from the diagonal of co-variance matrix of the fit to retention data. However, the resulting weights varied by 10 orders of magnitude and had to be bounded between normalized values of 0.56 and 1.8 for optimal PTF performance. Unbounded weights resulted in PTFs inferior to those by Schaap et al. (2001).
- The recalibrated PTFs have better performance in terms of root mean square error (*RMSE*) and mean error (*ME*) than the models in Schaap et al. (2001). For water retention, *ME* were improved from -0.022 (H1) to 0.0051 (H1w) cm^3/cm^3 and -0.013 (H5) to 0.0017 (H5w), indicating a substantial reduction in model bias. *ME* for K_s were almost unchanged as expected. *RMSE* for water retention were reduced modestly by about 5% (H1w) to 15% (H4w) compared to the original model, while *RMSE* for K_s is increased slightly by 0.9% (H3w) to 3.3% (H5w) in the new models on log scale of K_s compared with the Rosetta1 model.
- When plotted against pressure head (h), reduction of water retention in *ME* was present for $h > 32$ cm. Substantial positive and negative *ME* remained for $0 < h \leq 32$ cm, because the calibration database had relatively few samples with such pressure-heads. We speculate that the poor performance is due to macroporosity-driven changes to water content very close to saturation, but it

is impossible to resolve such issues with the current database.

- For each input soil sample, H2w through H5w estimate 1000 sets of soil hydraulic parameters (θ_r , θ_s , α , n , K_s). Univariate distributions of the estimated parameters are predominantly non-Gaussian and can be fitted rather well with α -stable distributions (Samoradnitsky and Taqqu, 1994) with parameters that are shift (μ), scale (σ), skewness (β), and the stability index (α). It was found that stability index α (in α -stable distribution) was < 2 , therefore corresponding distribution deviates from Gaussian distribution (the distribution is Gaussian distribution when $\alpha = 2$). This deviation was found especially for soil samples in the extreme silt and clay corners of the textural triangle that are not represented well in Rosetta's calibration database. However, the distributions of the estimates were found to be nearly symmetric.

- Despite the fact that the distribution of the estimated soil hydraulic parameters is more efficiently described by α -stable distribution, it was found that the arithmetic mean of the estimated parameters typically exhibits only a small bias. In addition, the numerical fit of α -stable distributions is slow and somewhat error-prone. We therefore recommend the continued use of arithmetic mean and (co-) variances (which implies $\alpha = 2$ and $\beta = 0$ in α -stable distribution) or even the use of median values (which approximate μ in α -stable distribution rather well). However, risk-based analyses may benefit from the use of α -stable characterizations, but could also be run directly on the estimated distributions by means of Monte-Carlo methods.

- The new models, as well as those by Schaap et al. (2001), were implemented in a Python script under an open-source license. A web-based user interface is being developed separate from the present work.

Acknowledgment

The work was supported, in part, by a grant from U.S. Department of Agriculture-Agriculture and Food Research Initiative (Grant No. 2014-67019-21718).

517

518 **References**

- 519 Assouline, S., Or, D., 2013. Conceptual and Parametric Representation of Soil Hydraulic Properties: A
520 Review. *Vadose Zo. J.* 12, 1–20. doi:10.2136/vzj2013.07.0121
- 521 Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble
522 Kalman filter. *Adv. Water Resour.* 29, 1107–1122.
- 523 Dai, Y., Shanguan, W., Duan, Q., Liu, B., Fu, S., Niu, G., 2013. Development of a China Dataset of
524 Soil Hydraulic Parameters Using Pedotransfer Functions for Land Surface Modeling. *J.*
525 *Hydrometeorol.* 14, 869–887. doi:10.1175/JHM-D-12-0149.1
- 526 Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- 527 Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2014. Anisotropic statistical scaling of soil and
528 sediment texture in a stratified deep vadose zone near Maricopa, Arizona. *Geoderma* 214–215,
529 217–227. doi:10.1016/j.geoderma.2013.09.008
- 530 Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2013. Anisotropic statistical scaling of vadose
531 zone hydraulic property estimates near Maricopa, Arizona. *Water Resour. Res.* 49, 8463–8479.
532 doi:10.1002/2013WR014286
- 533 Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning Data Mining,*
534 *Inference, and Prediction*, Springer. New York, New York.
- 535 Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. *J. Comput. Graph.*
536 *Stat.* 5, 299–314. doi:10.1080/10618600.1996.10474713
- 537 Li, Y., Chen, D., White, R.E., Zhu, A., Zhang, J., 2007. Estimating soil hydraulic properties of Fengqiu
538 County soils in the North China Plain using pedo-transfer functions. *Geoderma* 138, 261–271.
539 doi:10.1016/j.geoderma.2006.11.018
- 540 McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil
541 inference systems. *Geoderma* 109, 41–73. doi:10.1016/S0016-7061(02)00139-8

542 Minasny, B., McBratney, A., 2002. The Neuro-m Method for Fitting Neural Network Parametric
 543 Pedotransfer Functions. *Soil Sci. Soc. Am. J.* 66, 352–361.
 544 National Cooperative Soil Survey, 2016. National Cooperative Soil Survey Characterization Database.
 545 <http://ncsslabdatamart.sc.egov.usda.gov/> Accessed 2016-01-01.
 546 Neuman, S.P., 2005. Trends, prospects and challenges in quantifying flow and transport through
 547 fractured rocks. *Hydrogeol. J.* 13, 124–147. doi:10.1007/s10040-004-0397-2
 548 Neuman, S.P., 1982. Statistical characterization of aquifer heterogeneities: An overview. *GSA Spec.*
 549 *Pap.* 189, 81–102. doi:10.1130/SPE189-p81
 550 Nolan, J.P., 2001. Maximum likelihood estimation of stable parameters, in: Barndorff-Nielsen, O.E.,
 551 Mikosch, T., Resnick, S.I. (Eds.), *Levy Processes: Theory and Applications*. Birkhäuser Boston,
 552 pp. 379–400.
 553 Nolan, J.P., 1997. Numerical calculation of stable densities and distribution functions. *Commun. Stat.*
 554 *Stoch. Model.* 13, 759–774. doi:10.1080/15326349708807450
 555 Pachepsky, Y.A., Rawls, W.J., 2004. Development of pedotransfer functions in soil hydrology.
 556 *Developments in Soil Science*, Elsevier, Amsterdam.
 557 Pachepsky, Y.A., Rawls, W.J., 1999. Accuracy and Reliability of Pedotransfer Functions as Affected by
 558 Grouping Soils. *Soil Sci. Soc. Am. J.* 63, 1748. doi:10.2136/sssaj1999.6361748x
 559 Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. *Geoderma*
 560 131, 308–316. doi:10.1016/j.geoderma.2005.03.012
 561 Pachepsky, Y. a., Timlin, D., Varallyay, G., 1996. Artificial Neural Networks to Estimate Soil Water
 562 Retention from Easily Measurable Data. *Soil Sci. Soc. Am. J.* 60, 727.
 563 doi:10.2136/sssaj1996.03615995006000030007x
 564 Pachepsky, Y., Schaap, M.G., 2004. Data mining and exploration techniques. *Dev. Soil Sci.* 30, 21–32.
 565 doi:10.1016/S0166-2481(04)30002-4
 566 Panzeri, M., Riva, M., Guadagnini, A., Neuman, S.P., 2014. Comparison of Ensemble Kalman Filter

567 groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo
568 simulation. *Adv. Water Resour.* 66, 8–18.

569 Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *Numerical recipes in C: the art of scientific*
570 *computing*. Cambridge University Press.

571 Pringle, M.J., Romano, N., Minasny, B., Chirico, G.B., Lark, R.M., 2007. Spatial evaluation of
572 pedotransfer functions using wavelet analysis. *J. Hydrol.* 333, 182–198.
573 doi:10.1016/j.jhydrol.2006.08.007

574 Rawls, W.J., Brakensiek, D.L., 1985. Prediction of soil water properties for hydrologic modeling, in:
575 *Watershed Management in the Eighties*. American Society of Civil Engineers, pp. 293–299.

576 Rawls, W.J., Gish, T.J., Brakensiek, D.L., 1991. Estimating soil water retention from soil physical
577 properties and characteristics, in: *Advances in Soil Science*. Springer New York, pp. 213–234.
578 doi:10.1007/978-1-4612-3144-8_5

579 Romano, N., Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain
580 attributes. *J. Hydrol.* 265, 56–75. doi:10.1016/S0022-1694(02)00094-X

581 Samoradnitsky, G., Taqqu, M.S., 1994. *Stable non-Gaussian random processes: stochastic models with*
582 *infinite variance*. Chapman & Hall, New York.

583 Schaap, M.G., Bouten, W., 1996. Modeling water retention curves of sandy soils using neural
584 networks. *Water Resour. Res.* 32, 3033–3040. doi:10.1029/96WR02278

585 Schaap, M.G., Leij, F.J., 2000. Improved Prediction of Unsaturated Hydraulic Conductivity with the
586 Mualem-van Genuchten Model. *Soil Sci. Soc. Am. J.* 64, 843–851.
587 doi:10.2136/sssaj2000.643843x

588 Schaap, M.G., Leij, F.J., 1998a. Using neural networks to predict soil water retention and soil hydraulic
589 conductivity. *Soil Tillage Res.* 47, 37–42. doi:10.1016/S0167-1987(98)00070-1

590 Schaap, M.G., Leij, F.J., 1998b. Database-related accuracy and uncertainty of pedotransfer functions.
591 *Soil Sci.* 163, 765–779.

592 Schaap, M.G., Leij, F.J., van Genuchten, M.T., 2001. ROSETTA: A computer program for estimating
593 soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* 251, 163–176.
594 doi:10.1016/S0022-1694(01)00466-8

595 Schaap, M.G., Leij, F.J., van Genuchten, M.T., 1998. Neural network analysis for hierarchical
596 prediction of soil hydraulic properties. *Soil Sci. Soc. Am. J.* 62, 847–855.
597 doi:10.2136/sssaj1998.03615995006200040001x

598 Schaap, M.G., Nemes, A., van Genuchten, M.T., 2004. Comparison of Models for Indirect Estimation
599 of Water Retention and Available Water in Surface Soils. *Vadose Zo. J.* 3, 1455–1463.
600 doi:10.2136/vzj2004.1455

601 Schweppe, F.C., 1973. Uncertain dynamic systems. Prentice Hall, Inc., Englewood Cliffs, New Jersey.

602 Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples).
603 *Biometrika* 52, 591–611. doi:10.2307/2333709

604 Shi, L., Song, X., Tong, J., Zhu, Y., Zhang, Q., 2015. Impacts of different types of measurements on
605 estimating unsaturated flow parameters. *J. Hydrol.* 524, 549–561.

606 Šimůnek, J., van Genuchten, M.T., Šejna, M., 2012. The HYDRUS software package for simulating
607 two and three-dimensional movement of water, heat, and multiple solutes in variably-saturated
608 media: Technical manual, version 1.0, technical manual, PC Progress. Prague, Czech Republic.

609 Šimůnek, J., van Genuchten, M.T., Šejna, M., 2008. Development and Applications of the HYDRUS
610 and STANMOD Software Packages and Related Codes. *Vadose Zo. J.* 7, 587.
611 doi:10.2136/vzj2007.0077

612 Song, X., Shi, L., Ye, M., Yang, J., Navon, I.M., 2014. Numerical comparison of iterative ensemble
613 Kalman filters for unsaturated flow inverse modeling. *Vadose Zo. J.* 13.

614 Sudicky, E.A., 1986. A natural gradient experiment on solute transport in a sand aquifer: Spatial
615 variability of hydraulic conductivity and its role in the dispersion process. *Water Resour. Res.* 22,
616 2069–2082. doi:10.1029/WR022i013p02069

617 Twarakavi, N.K.C., Šimůnek, J., Schaap, M.G., 2009. Development of Pedotransfer Functions for
618 Estimation of Soil Hydraulic Parameters using Support Vector Machines. *Soil Sci. Soc. Am. J.* 73,
619 1443–1452. doi:10.2136/sssaj2008.0021

620 van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of
621 Unsaturated Soils. *Soil Sci. Soc. Am. J.* 44, 892–898.

622 Venables, W.N., Smith, D.M., 2003. An Introduction to R, The R Development Core Team.

623 Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., Genuchten, M.T. Van, 2010.
624 Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties:
625 A Review. *Vadose Zo. J.* 9, 795. doi:10.2136/vzj2010.0045

626 Vrugt, J.A., Stauffer, P.H., Wöhling, T., Robinson, B.A., Vesselinov, V. V., 2008. Inverse Modeling of
627 Subsurface Flow and Transport Properties: A Review with New Developments. *Vadose Zo. J.* 7,
628 843–864. doi:10.2136/vzj2007.0078

629 Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of
630 hydraulic properties of European soils. *Geoderma* 90, 169–185. doi:10.1016/S0016-
631 7061(98)00132-3

632 Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: Bridging the gap between
633 available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 251, 123–150.
634 doi:10.1016/S0022-1694(01)00464-4

635 Zhang, Y., Schaap, M.G., Guadagnini, A., Neuman, S.P., 2016. Inverse Modeling of Unsaturated Flow
636 Using Clusters of Soil Texture and Pedotransfer Functions. *Water Resour. Res.* 52, 1–14.
637 doi:10.1002/2016WR019016

640

641 Figure 1. Sorted cumulative contribution of weights for VG parameters versus number of samples. One single sample with
642 the largest weight was removed for θ_s and α in Figure 1b and 1c for display reasons; the weight of this single sample was
643 orders of magnitude larger than others.

Figure 2. $RMSE_w$ and ME_w values for different weighted water retention models (a) H2w, (b) H3w, (c) H4w, (d) H5w models.

Figure 3. $RMSE_w$ (left axis) and $RMSE_{K_s}$ (right axis) values of H3w in Rosetta3 models vs different weights for K_s ($\Omega_U=1.8$ and $\Omega_L=1/1.8$ for the weights of water retention parameters).

Figure 4. $RMSE_w$ (left axis) and $RMSE_{Ks}$ (right axis) values of H3w in Rosetta3 models vs different hidden nodes ($\Omega_U=1.8$ and $\Omega_L=1/1.8$ for the weights of water retention parameters and eight for the weights of K_s).

Figure 5. (a) $RMSE_w$, (b) $RMSE_{Ks}$, (c) ME_w , and (d) ME_{Ks} vs different number of bootstrap of H3w in Rosetta3 model.

Figure 6. ME_w (lines, left axis) for different pressure head in (a-e) H1 through H5 in Rosetta1 model and (f-j) H1w through H5w in Rosetta3 models. ME_w values were computed for nine pressure head classes with boundaries at 0, 3.2, 10, 32, 100, 320, 1000, 3200, 10 000, and 32 000 cm. For reference, the direct fit to the original retention data is shown (FIT). The number of retention points for each suction class is also shown (bars, right axis).

657

658 Figure 7. (a) Soil textural distribution of Rosetta's database, including 2134 soil samples for water retention and 1306
 659 samples for K_s ; (b-f) p -value distribution of Shapiro-Wilk test of VG parameters and K_s estimated from H2w in Rosetta3
 660 model using 1000 bootstrap realizations; the results are estimated on every whole percentage of the soil textural triangle. S,
 661 sand; LS, loamy sand; sL, sandy loam; sCL, sandy clay loam; sC, sandy clay; L, loam; siL, silty loam; Si, silt; siCL, silty clay
 662 loam; siC, silty clay; cL, clay loam; C, clay.
 663

664

665

666

667

668

669

670

671

672

Figure 8. Diagonal graphs: histograms of VG parameters and K_s estimated from H2w in Rosetta3 model using 1000 bootstrap realizations for an illustrated soil sample with sand, silt, and clay percentage of 40%, 20%, and 40%. $RMSE$ with G and α subscript indicates Gaussian distribution and α -stable distribution, respectively. Lower triangle graphs: scatter points of corresponding VG parameters and K_s pairs. The center points calculated at mean values, median values, μ values, and 95% confidence ellipse are also shown.

Figure 9. Diagonal graphs: histograms of VG parameters and K_s estimated from H2w in Rosetta3 model using 1000 bootstrap realizations for an illustrated soil sample with sand, silt, and clay percentage of 5%, 5%, and 90%. $RMSE$ with G and α subscript indicates Gaussian distribution and α -stable distribution, respectively. Lower triangle graphs: scatter points of corresponding VG parameters and K_s pairs. The center points calculated at mean values, median values, μ values, and 95% confidence ellipse are also shown.

Figure 10. The distribution of α value (in α -stable distribution) for VG parameters and K_s estimated from H2w in Rosetta3 model using 1000 bootstrap realizations; the results are estimated on every whole percentage of the textural triangle.

Figure 11. The distribution of $(\text{mean}-\mu)/\text{mean}$ value for VG parameters and K_s estimated from H2w in Rosetta3 model using 1000 bootstrap realizations; the results are estimated on every whole percentage of the soil textural triangle. Mean value is arithmetic mean of estimated VG and K_s ; μ value is estimated from α -stable distribution.

689 **Tables**

690 Table 1. Ratios between estimates of H1 (Rosetta1, Schaap et al., 2001) and H1w models (Rosetta3, this study) for the
691 twelve USDA soil textural classes. Each pairs of columns give the ratios of the means and the ratios of the standard
692 deviations (abbreviated as sd); the ratio of sd for K_s is omitted (equal to 1 for all texture classes). The “soiltexture” package
693 (<https://cran.r-project.org/web/packages/soiltexture/index.html>) in R is used to determine the soil texture classes. Several
694 samples are located at the boundary of two classes and are therefore included in both classes. Consequently, the total
695 number of soil samples in texture classes is larger than 2134.
696
697

Soil texture classes	θ_r (cm ³ cm ⁻³)		θ_s (cm ³ cm ⁻³)		α (1/cm)		N		K_s (cm/day)	Number of samples
	mean	sd	mean	sd	mean	sd	mean	sd	mean	
Clay	1.33	0.99	1.00	1.07	0.57	0.93	1.00	1.00	1.00	92
Silty Clay	1.12	0.98	0.98	1.12	0.62	1.00	0.96	0.85	1.00	29
Sandy clay	1.26	0.84	0.99	1.02	0.75	0.90	1.02	1.03	1.00	12
Clay loam	1.32	0.96	0.98	1.06	0.59	0.94	0.99	0.85	1.00	147
Silty clay loam	1.28	0.91	0.98	1.11	0.64	0.86	0.94	0.82	1.00	92
Sandy clay loam	1.52	1.03	0.99	0.90	0.58	0.88	0.99	0.91	1.00	179
Loam	1.44	1.02	1.00	0.96	0.59	0.88	0.96	0.86	1.00	262
Silty loam	1.30	0.96	0.97	0.89	0.68	0.83	0.93	0.89	1.00	332
Sandy loam	1.53	1.11	0.98	0.88	0.62	1.01	1.00	0.98	1.00	487
Silt	1.30	0.8	0.97	1.06	0.92	0.94	0.94	0.83	1.00	6
Loamy sand	1.20	0.96	0.98	0.91	0.71	0.97	0.97	0.90	1.00	205
Sand	1.03	0.97	0.97	0.87	0.93	0.97	0.91	0.95	1.00	308

698
699

700
701

Table 2. ME_w , $RMSE_w$, ME_{K_S} , and $RMSE_{K_S}$ evaluated for calibration (C) and validation (V) data of Rosetta3 models.

	$ME_w(C)$	$ME_w(V)$	$RMSE_w(C)$	$RMSE_w(V)$	$ME_{K_S}(C)$	$ME_{K_S}(V)$	$RMSE_{K_S}(C)$	$RMSE_{K_S}(V)$
H2w	0.0042	0.0042	0.072	0.072	0.002	0.002	0.723	0.733
H3w	0.0063	0.0062	0.061	0.062	0.001	0.001	0.667	0.681
H4w	0.0023	0.0022	0.040	0.040	0.004	0.005	0.594	0.610
H5w	0.0018	0.0017	0.039	0.039	0.004	0.004	0.593	0.612

702

703 Table 3. ME_w , $RMSE_w$, ME_{Ks} , and $RMSE_{Ks}$ evaluated for Rosetta1 and Rosetta3 models.
704

	a: Rosetta1				b: Rosetta3			
	ME_w	$RMSE_w$	ME_{Ks}	$RMSE_{Ks}$	ME_w	$RMSE_w$	ME_{Ks}	$RMSE_{Ks}$
H1	-0.022	0.078	-0.001	0.739	0.0051	0.074	0.001	0.739
H2	-0.021	0.076	-0.001	0.717	0.0042	0.072	0.002	0.726
H3	-0.022	0.068	0.000	0.666	0.0063	0.062	0.001	0.672
H4	-0.019	0.047	-0.004	0.586	0.0022	0.040	0.004	0.600
H5	-0.013	0.044	-0.002	0.581	0.0017	0.039	0.004	0.600

705

706 Table 4. R² between calibration (C) results of soil hydraulic parameters estimated from Rosetta3 and fitted (or observed)
 707 parameters. Validation results are also shown for comparison.

	θ_r		θ_s		$\log_{10}(\alpha)$		$\log_{10}(n)$		$\log_{10}(Ks)$	
	C	V	C	V	C	V	C	V	C	V
H2w	0.147	0.136	0.190	0.177	0.324	0.315	0.460	0.453	0.453	0.437
H3w	0.147	0.134	0.574	0.564	0.324	0.314	0.473	0.464	0.534	0.514
H4w	0.162	0.146	0.604	0.594	0.489	0.479	0.565	0.554	0.630	0.610
H5w	0.471	0.451	0.612	0.602	0.601	0.591	0.768	0.761	0.632	0.608

708
 709

710 Table 5. R^2 between fitted / observed parameters and parameters estimated from Rosetta3 models. For comparison, R^2
 711 between fitted / observed parameters and parameters estimated from Rosetta1 is also shown.
 712

	θ_r		θ_s		$\log_{10}(\alpha)$		$\log_{10}(n)$		$\log_{10}(Ks)$	
	Rosetta1	Rosetta3	Rosetta1	Rosetta3	Rosetta1	Rosetta3	Rosetta1	Rosetta3	Rosetta1	Rosetta3
H1	0.066	0.116	0.143	0.154	0.203	0.305	0.452	0.441	0.427	0.426
H2	0.086	0.143	0.178	0.186	0.238	0.321	0.473	0.457	0.461	0.447
H3	0.094	0.142	0.581	0.570	0.265	0.320	0.495	0.470	0.535	0.527
H4	0.121	0.156	0.605	0.600	0.417	0.485	0.599	0.561	0.640	0.623
H5	0.387	0.464	0.600	0.609	0.577	0.598	0.760	0.766	0.647	0.623

713