Regularized Artificial Neural Network Training for Biased Data of Soil Hydraulic Parameters

Mahesh R. Gautam,[1] Jianting Zhu,[1] and Ming Ye[2]

Abstract: Development and application of artificial neural network (ANN) pedotransfer functions for estimating soil hydraulic properties (SHP) have become popular in the last two decades. However, limited availability of SHP training data often constrains the full potential of improved SHP estimation with ANN in many practical situations. In many situations, SHP data are limited and could be biased by samples from a restricted portion of the data population. Artificial neural network pedotransfer functions developed under such situations are likely to yield biased estimates. We proposed a direct approach to minimize mean estimation errors (bias) in such situations and developed a regularized ANN algorithm. The new algorithm revised the ANN error function and its gradients with respect to neural network outputs. We applied the new algorithm to synthetically generated SHP data representing different data availability situations and found that the newly developed algorithms were effective in reducing bias. Training with both the new and conventional mean square error functions resulted in equally good results in test phases when ANN models were trained with randomly sampled unbiased data. However, when ANN was trained with and applied to SHP data with respectively different means (biased sample), the proposed regularized ANN was highly effective in minimizing the bias when compared with ANN with the conventional mean square error function.

Key words: Artificial neural network, mean square error, regularized training, revised error function.

(*Soil Sci* 2011;176: 00–00)

The soil hydraulic property (SHP) pedotransfer functions (hereafter called pedotransfer function or PTF) relate SHP to readily available soil texture and other data and are a viable, practical approach for the estimation of SHP (Rawls and Brankensiek, 1985; van Genuchten and Leij, 1992; Koekkoek and Bootlink, 1999; Wösten et al., 2001). Although earlier PTF were essentially based on statistical linear regression (e.g., Rawls and Brankensiek, 1985; Saxton et al., 1986 etc.), artificial neural network (ANN)–based nonlinear regressions are on a rapid rise since the mid-1990s (e.g., Pachepsky et al., 1996; Schaap and Bouten, 1996; Tamari et al., 1996; Schaap et al., 1998; Schaap and Leij, 1998; Minasny et al., 1999; Schaap et al., 2001; Minasny and McBratney, 2002; Nemes et al., 2003; Børgesen and Schaap, 2005; Merdun et al., 2006; Ye et al., 2007; Stumpp et al., 2009; among others). This surge in ANN application stems from the ability of ANN to capture potential nonlinear relationships between SHP and their predictors (soil texture and other properties; Schaap et al., 2001), the lack of required *a priori* models in ANN modeling (Pachepsky and Schaap, 2004), and better predictive power (Minasny, 2007). Despite such advantages, the ANN PTF, like its other statistical counterparts, faces a serious practical constraint in its applicability because of limited availability of SHP data.

For accurate prediction, ANN requires a large SHP data set for training and validation. In many practical situations, SHP data are limited, and available SHP data could be biased samples from a restricted portion of the data population. The ANN model trained (calibrated) and validated on such data would result in biased predictions.

Reported approaches to bias correction use an indirect, two-stage scheme (Jana et al., 2007; Jana et al., 2008) in which ANN is trained without considering potential future (test area) data, and then bias correction is applied to ANN model output using statistical techniques such as bias correction and cumulative frequency distribution mapping (Ines and Hansen, 2006). In this study, we present an alternative, direct approach in which the bias correction process is embedded in the ANN algorithm itself under the premise that more-representative hydraulic parameters means are available or can be determined for the area in which PTF has to be developed. However, as demonstrated later (see An Illustrative Example), our algorithm is applicable to both cases: with and without bias.

In many instances, hydraulic properties used in ANN development are determined from soil samples that may not be representative of an entire population for which the model is intended; however, mean SHP may be estimated based on soil type or from expert elicitation. Several compiled databases (UNSODA, HYPRES, WISE, USDA; Wösten et al., 2001) are available for soil texture and SHP in different regions of the world. A reasonable prior mean for any study area may be obtained from these regional databases. There are other occasions where mean hydraulic properties may be derived from upscaling algorithms for large areas based on remote sensing of soil moisture. But as SHP data are determined in a specific zone such as those used in the ANN development, the data can be used to update the prior mean to obtain the target mean required in the ANN development proposed in this study. This target mean might be different from either the sample mean estimated by the neural network or the prior mean, but we believe is more representative of the actual mean and could be used as the target mean in ANN algorithm development. Although we did not explore such updating in our study, the Bayesian updating approach (e.g., Meyer et al., 1997; Ye et al., 2007; Woodbury, 1989) shows such a possibility.

This study is intended to make full use of available information in developing ANN models of SHP estimates, by integrating measured SHP with prior information on mean SHP. The main contribution of the proposed method is that it can incorporate a reasonable target mean into ANN development in a formal way. In this study, we focus mainly on developing the new ANN training algorithm. To the best of our knowledge, such a training algorithm has not been reported in soil literature before.

[1]Division of Hydrologic Sciences, Desert Research Institute, Nevada System of Higher Education, Las Vegas, NV 89119. Dr. Mahesh R. Gautam is corresponding author. E-mail: mahesh.gautam@dri.edu
[2]Department of Scientific Computing, Florida State University, Tallahassee, FL.
Received March 25, 2011.
Accepted for publication August 4, 2011.
Financial Disclosures/Conflicts of Interest: *This study was financially supported by the U.S. Department of Energy under grant DE-FG02-06ER46265, the U.S. Geological Survey grant 06HQGR0098, the Applied Research Initiative of Nevada, and Maki Chair Program at Desert Research Institute.*

ISSN: 0038-075X
DOI: 10.1097/SS.0b013e3182316c93

It should be noted that the proposed neural network algorithm needs the same input and output variables as the conventional one, but uses a revised error function (EF) (objective function) that includes a penalty or regularization term in addition to the mean square error (MSE), a widely used ANN error minimization function (see New Approach in Neural Network Model Development). Our proposed approach imposes a penalty on calculated statistics (e.g., mean) used in network estimation so that deviation from known statistics is penalized. Since new forms of EF are used, new gradients of the EF with respect to the neural network weights are required. In this article, we present our new approach and formulation and demonstrate the effectiveness through an illustrative ANN experiment with synthetic data. Although only the ANN development for correction of target mean is presented in this article, the development method can be extended for correction of higher moments such as variance, skewness, and kurtosis.

MATERIALS AND METHODS

The study was designed with a synthetic soil texture data set and resulting SHP, discussed in detail in An Illustrative Example. A brief description of the SHP, derivation of new training algorithm, and ANN modeling approach are discussed in this section.

Hydraulic Properties

We used the van Genuchten (1980) model and Mualem (1976) model to represent SHP. The van Genuchten (1980) model for the soil-water retention curve combined with the hydraulic conductivity function (Mualem, 1976; Vereecken et al., 2010) can be expressed as follows:

$$\Theta(h) = \Theta_r + \left(\Theta_s - \Theta_r\right)\left[1 + (\alpha h)^n\right]^{-m} \quad (1)$$

$$K(h) = K_s\left\{1 - (\alpha h)^{mn}\left[1 + (\alpha h)^n\right]^{-m}\right\}^2\left[1 + (\alpha h)^n\right]^{-m/2} \quad (2)$$

where h is the soil-water suction head [cm]; Θ is the soil-water content ($cm^3\ cm^{-3}$); Θ_r is the residual water content ($cm^3\ cm^{-3}$); Θ_s is the saturated water content ($cm^3\ cm^{-3}$); K_s is the saturated hydraulic conductivity ($cm\ h^{-1}$); α is the shape factor, approximately equal to the inverse of the air entry value (cm^{-1}); n is the pore size distribution index [−]; and m is the empirical constant, which can be related to n by $m = 1 - 1/n$.

New Approach in Neural Network Model Development

Artificial neural network training involves finding optimal weights in the network mapping inputs to outputs. For this purpose, conventional neural network–based PTF utilize the MSE EF (Eq. 3), which is the weighted sum of squared errors of neural network predicted soil hydraulic parameters.

$$\text{MSE EF} = \sum_{i=1}^{N_s}\sum_{j=1}^{N_o}\beta_{ij}(\hat{Y}_{ij} - Y_{ij})^2 \quad (3)$$

where N_s is number of samples; N_o is number of output parameters (i.e., $\log_{10}K_s$, Θ_r, $\log_{10}\alpha$, and $\log_{10}n$); Y_{ij} is the measured value of the j-th hydraulic parameter of the i-th sample, $\hat{Y}_{ij}$ is the corresponding predicted hydraulic parameter based on the ANN model; and β_{ij} is a weighting coefficient assigned based on the confidence level of Y_{ij} that is not adjusted in the neural network training process. In this study, we developed new ANN models that incorporate statistical properties, such as mean hydraulic parameter data, by modifying the MSE as follows. The second term in Eq. 4 is expected to result in bias correction: the correction of the estimated mean from the actual one (target mean). Thus, for the bias correction, we use the following revised EF (REF, Eq. 4).

$$\text{REF} = \sum_{i=1}^{N_s}\sum_{j=1}^{N_o}\beta_{ij}(\hat{Y}_{ij} - Y_{ij})^2 + \sum_{j=i}^{N_o}\gamma_j(\hat{\mu}_j - \mu_j)^2 \quad (4)$$

where, γ_j is a preassigned weighting coefficient for the mean; $\hat{\mu}_j$ and μ_j, respectively, are the mean of SHP estimated by neural networks and the target mean, assumed to be known *a priori* from processes mentioned earlier. As we discussed earlier, REF may be more useful than the conventional MSE in practical applications where bias of SHP data may occur when the hydraulic parameter data were sampled in only part of the SHP population. Artificial neural network estimations based on the conventional EF (MSE) will introduce biased statistics of the predicted variables. The proposed REF (Eq. 4) penalizes differences between the ANN estimated mean and the *a priori* known target mean value.

With the nonconventional EF, a new neural network training algorithm is needed. Computation of derivatives of the EF with respect to the neural network weights is a key step in implementation of neural network models for derivative-based error minimization. In MATLAB's Neural Network Toolbox, this can be accomplished by redefining the EF as in Eq. 4 and then deriving the derivative of the EF with respect to output hydraulic parameters. The derivations of the derivatives are shown below. The derivative of the first term in REF (Eq. 4) with respect to the output hydraulic parameter is as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\hat{Y}_{kl}}\sum_{i=1}^{N_s}\sum_{j=1}^{N_o}\beta_{ij}(\hat{Y}_{ij} - Y_{ij})^2 \\
&\quad= \sum_{i=1}^{N_s}\sum_{j=1}^{N_o}\frac{\partial}{\partial\hat{Y}_{kl}}\{\beta_{ij}(\hat{Y}_{ij} - Y_{ij})^2\} \\
&\quad= \sum_{i=1}^{N_s}\sum_{j=1}^{N_o}2\beta_{ij}(\hat{Y}_{ij} - Y_{ij})\left(\frac{\partial\hat{Y}_{ij}}{\partial\hat{Y}_{kl}}\right) \\
&\quad= \sum_{i=1}^{N_s}\sum_{j=1}^{N_o}2\beta ij(\hat{Y}_{ij} - Y_{ij})\delta_{ij}\delta_{jl} \\
&\quad= 2\beta_{kl}(\hat{Y}_{kl} - Y_{kl})
\end{aligned} \quad (5)$$

Derivative of the second term in REF (Eq. 4) with respect to the output hydraulic parameter is as follows:

$$\begin{aligned}
&\frac{\partial}{\partial\hat{Y}_{kl}}\sum_{j=1}^{N_o}\gamma_j(\hat{\mu}_j - \mu_j)^2 \\
&\quad= \sum_{j=1}^{N_o}\frac{\partial}{\partial\hat{Y}_{kl}}\{(\gamma_j(\hat{\mu}_j - \mu_j)^2\} \\
&\quad= \sum_{j=1}^{N_o}2\gamma_j(\hat{\mu} - \mu_j)\left(\frac{\partial\hat{\mu}_j}{\partial\hat{Y}_{kl}}\right) \\
&\quad= \sum_{j=1}^{N_o}2\gamma_j(\hat{\mu}_j - \mu_j)\frac{\partial}{\partial\hat{Y}_{kl}}\left(\frac{1}{N_s}\sum_{m=1}^{N_s}\hat{Y}_{jm}\right) \\
&\quad= \sum_{j=1}^{N_o}2\gamma_j(\hat{\mu}_j - \mu_j)\frac{1}{N_s}\frac{\partial}{\partial\hat{Y}_{kl}}\left(\sum_{m=1}^{N_s}\hat{Y}_{jm}\right) \\
&\quad= \sum_{j=1}^{N_o}2\gamma_j(\hat{\mu}_j - \mu_j)\frac{1}{N_s}\delta_{jk} \\
&\quad= \frac{2}{N_s}\gamma_k(\hat{\mu}_k - \mu_k)
\end{aligned} \quad (6)$$

Where δ_{ik}, δ_{jl} are the Kronecker deltas, with properties $\delta_{ik} = 1$ for $i = k$ and $\delta_{jl} = 1$ for $j = l$ and zero elsewhere.

The chosen EF (MSE or REF) were implemented by specifying them as parameters of the "newff" function in

a MATLAB script written to implement the feed-forward back-propagation ANN. For the regularized ANN, a revised MSE ("REF function") had to be implemented rather than MSE. The mean of the min-max–scaled values of the output of the test phase (assumed to be known *a priori*) was specified in the REF function that implemented Eqs. 5 and 6. Values for weighting coefficients β and γ were specified within the REF function.

It should be noted that, although not implemented here, the approach can be extended to include bias corrections of other moments (e.g., S.D., skewness, and kurtosis) by introducing suitable regularization terms in the EF and similarly finding derivatives of the EF with respect to the neural network weights. The mathematics, however, is more sophisticated, and more numerical experiments are required. Practically, the information on these higher-order moments required for the ANN training would be much more difficult to obtain than for the mean. The current form of ANN implementation with only a penalty term resembles linear bias correction as in Jana et al. (2007), but the advantage of this approach is that bias correction can be achieved through an ANN training in a single-stage process.

Modeling Approach

A back-propagation type, two-layer feed-forward ANN (Bishop, 1995) with an input layer (with N_I nodes corresponding to the number of input variables *viz.* S, C, Θ_s; see An Illustrative Example), a hidden layer (with K nodes), and an output layer (with N_o nodes corresponding to output variables) was adopted in the study. Outputs considered in our study consisted of all four hydraulic parameters, namely, $\log_{10}K_s$, Θ_r, $\log_{10}\alpha$, and $\log_{10}n$ (see An Illustrative Example). A single-layer network has been recommended and is used in practice in both surface and sub-surface hydrology, including ANN PTF cases, because a single layer is considered sufficient for generalization (Hornik et al., 1989). Although a number of heuristics are available (Maier and Dandy, 2000), the determination of an optimal number of hidden nodes is essentially a trial-and-error process. Although a small number of hidden nodes cannot generalize, a larger number of hidden nodes can be computationally burdensome and may overfit, particularly if no independent validation data set is available. For a data set of moderate size like that in our study, and with a rough intuition gained through some quick, preliminary experiments, we adopted six hidden nodes for the hidden layer. As ANN PTF application deals with a prediction-type problem, we used a hyperbolic tangent for nodes in the hidden layer and a linear function for nodes in the output layer (Masters, 1993). We adopted an early stopping approach (Bishop, 1995) to avoid overfitting. The training data sets were divided into two groups: 75% of the data set was allocated for training, and 25% was allocated for validation. Training stopped when the validation data sets started showing an increase in error level. The ANN experiments were carried out in MATLAB's Neural Network Toolbox (Demuth and Beale, 1992).

Training for the MSE EF is straightforward when no prior information about mean and other statistics is available. However, training for REF that accounts for prior information is complicated and poses some difficulty. The conjugate gradient optimization algorithm commonly used in ANN training process is a local optimization approach, and error descent in the training is highly dependent on the initial starting point in the error surface. This, in turn, depends on the initial weight generated. Considering these complications, we applied 10 random weight generations for each ANN training and retained the best result for the training phase. As penalty terms are included in the REF to get better results in the test phase, training allows ANN results in the training phase to be different from the actual output. Because the test phase data are presumed to be unavailable, training is stopped based on training data (cross-validation error) or maximum number of epochs, whichever is reached earlier. The weighting coefficients β and γ in the REF (Eq. 4) are determined based on trial and error. The REF cannot capture variability of the output if a high weight (γ) is applied to the mean difference only. In other words, a large weight on the bias term would mean that variability of the output would not be preserved even if the simulated mean might come closer to the target one. On the other hand, if a very low weight is applied to the mean difference, bias correction would not be effective. Thus, a balance is required between these two terms. Without an exhaustive search for these coefficients, we could fix them after a few sample trials. A relatively large weight of 0.7 is used for the MSE part of the REF to preserve overall variability of the predicted parameters. For consistency, we used weighting coefficients β = 0.7 and γ = 0.3 in (Eq. 4) for all analyses. Note that these weighting coefficients are imposed on the statistics derived from the min-max–scaled values of the output.

AN ILLUSTRATIVE EXAMPLE

To evaluate the proposed approach, a numerical experiment was carried out using a synthetically generated data set. We used a statistical multiple regression equation developed by Rawls and Brankensiek (1985) and Carsel and Parrish (1988) to generate synthetic data for this purpose. The input data required to estimate hydraulic parameters are saturated water content (Θ_s in $cm^3\ cm^{-3}$), sand content (S in percentage), and clay content (C in percentage). The output parameters (i.e., the hydraulic parameters to be estimated) are the van Genuchten parameters Θ_r, α, n, and K_s. The form of general regression model (Carsel and Parrish, 1988) is as follows,

$$\begin{aligned} f(S,C,\Theta_s) = {} & b_0 + b_1 S + b_2 C + b_3 \Theta_S + b_{11} S^2 + b_{22} C^2 \\ & + b_{33}\Theta_S^2 + b_{12} SC + b_{13} S\Theta_S + b_{23} C\Theta_S \\ & + b_{112} S^2 C + b_{223} C^2 \Theta_S + b_{113} S^2 \Theta_S \\ & + b_{112} SC^2 + b_{233} C\Theta_S^2 + b_{1133} S^2 \Theta_S^2 \\ & + b_{2233} + C^2 \Theta_S^2 \end{aligned} \quad (7)$$

where f denotes any of the variables $\ln(K_s)$, Θ_r, $\ln(\alpha^{-1})$, or $\ln(n^{-1})$.

TABLE 1. Mean of Outputs in Training and Test Phases

	Training Phase				Test Phase			
	Θ_r	$\log_{10}K_s$	$\log_{10}n$	$\log_{10}\alpha$	Θ_r	$\log_{10}K_s$	$\log_{10}n$	$\log_{10}\alpha$
Case 1	0.042	−0.573	0.128	−1.507	0.072	0.666	0.153	−1.026
Case 2	0.085	1.129	0.170	−0.827	0.058	0.099	0.139	−1.252
Case 3	0.067	0.435	0.145	−1.125	0.065	0.278	0.149	−1.167
Case 4 (random)	0.066	0.346	0.147	−1.164	0.064	0.362	0.147	−1.137

TABLE 2. Difference (in Percentage) of Mean of Outputs in Test Phase Data Set Compared With Training Phase Data Set

	Θ_r	$\text{Log}_{10}K_s$	$\text{Log}_{10}n$	$\text{Log}_{10}\alpha$
Case 1	71.43	216.23	19.53	31.92
Case 2	31.76	91.23	18.23	51.40
Case 3	2.98	36.09	2.76	3.73
Case 4	3.03	4.51	0.000	2.32

We considered a case of sandy loam soil with the following distribution characteristics (mean and S.D.) from Carsel and Parrish (1988): $\bar{S}$ = 63.4%, σ_S = 7.9%; $\bar{C}$ = 11.1%, σ_C = 4.8%; $\bar{\Theta}_s$ = 0.41, $\sigma_{\Theta s}$ = 0.09. Using these statistics for the input parameters, we generated 300 data points for S, C, and Θ_s based on normal distributions using the Latin hypercube sampling method (McKay et al., 1979).

Saturated hydraulic conductivity and water retention parameters for the van Genuchten model were derived using the multiple-regression model (Eq.7) based on generated S, C, and Θ_s. The obtained parameters were converted into commonly used forms: $\log_{10}K_s$, Θ_r, $\log_{10}\alpha$, and $\log_{10}n$. We also generated white noise for each parameter with mean of zero and S.D. equal to 10% of the parameter's S.D. and added output from Eq. 7 as measurement error.

Although our focus here is on development of an algorithm, we must account for the different cases of bias (see paragraph 2). Therefore, we applied our algorithm to hypothetical cases of biases as considered in the following cases.

(i) Biased data (biased sampling)

- Case 1: Data from the first quartile is used for training (n = 75) and from other quartiles (n = 225) for testing.
- Case 2: Data from the last quartile (n = 75) is used for training and from other quartiles (n = 225) for testing.
- Case 3: Data from an inter-quartile range (n = 150) is used for training, while assigning remaining data (n = 150) for testing.

Note that Cases 1 and 2 represent extreme biased examples.

(ii) Unbiased data (random sampling)

- Case 4: 100 data points were randomly picked from the whole distribution (300 data points) for training and rest (200 data points) for testing.

Output target means from both biased and random data used for training and testing are shown in Table 1. The absolute difference of target means in the test phase as compared with the training phase (in percentage) is given in Table 2. As can be seen, there are distinct differences in the target mean for the biased data sets, particularly for Cases 1 and 2, for all parameters. Interestingly, unlike other parameters, $\log_{10}K_s$ also showed considerable difference in target mean between training and test phase data (about 36%) in Case 3. A closer look at the probability distribution of the parameter revealed that it had longer tail to the left compared with the other parameters. A D'Agostino-Pearson's K^2 test (D'Agostino et al., 1990) showed that $\log_{10}K_s$ data, and the other parameters, did not pass the normality test at 5% significance level. The inputs of the regression equation (Eq. 7) were normally distributed, so the regression equation accounts for the distribution of the parameters. A nonparametric Wilcoxon rank sum test was applied to the training and test data. The test showed that all parameters showed statistically significant (at 5% significance level) mean differences for Cases 1 and 2. For Case 3, again only $\log_{10}K_s$ showed statistically significant

TABLE 3. Performance of MSE and REF in Training and Test Phases in Biased and Random Sampling Cases for Soil Hydraulic Parameters (i) Θ_r, (ii) $\log_{10}K_s$, (iii) $\log_{10}n$, and (iv) $\log_{10}\alpha$

	R^2				Normalized Bias			
	Training		Test		Training		Test	
	MSE	REF	MSE	REF	MSE	REF	MSE	REF
(i) Θ_r								
Case 1	0.996	0.998	0.688	0.885	0.000	0.198	0.164	0.079
Case 2	0.997	0.995	0.770	0.865	0.000	0.095	0.117	0.059
Case 3	0.999	0.998	0.966	0.963	0.000	0.010	0.009	0.019
Case 4	1.000	0.997	0.998	0.974	0.000	0.007	0.002	0.002
(ii) $\text{Log}_{10}K_s$								
Case 1	1.000	0.986	0.719	0.912	0.004	0.657	0.637	0.097
Case 2	0.998	0.991	0.731	0.775	0.001	0.275	4.765	1.489
Case 3	0.998	1.000	0.929	0.966	0.002	0.108	0.340	0.037
Case 4	0.999	0.998	0.994	0.987	0.007	0.016	0.002	0.027
(iii) $\text{Log}_{10}n$								
Case 1	0.997	0.990	0.628	0.656	0.000	0.062	0.073	0.031
Case 2	0.996	0.996	0.646	0.748	0.000	0.055	0.043	0.026
Case 3	0.993	0.995	0.928	0.944	0.000	0.008	0.022	0.014
Case 4	0.964	0.975	0.969	0.971	0.003	0.011	0.001	0.013
(iv) $\text{Log}_{10}\alpha$								
Case 1	0.999	0.991	0.685	0.821	0.000	0.096	0.143	0.055
Case 2	0.997	0.983	0.440	0.715	0.000	0.155	0.082	0.018
Case 3	0.993	0.991	0.951	0.947	0.000	0.011	0.040	0.009
Case 4	0.999	0.995	0.983	0.970	0.000	0.010	0.000	0.010

difference in the mean. The other parameters in Case 3, and all parameters in Case 4, did not have statistically significant differences in target means between training and test phases.

Prediction error can be generally disaggregated into systematic and random errors. Thus, the goodness of fit is expressed with performance metrics that represent both errors. In this study, we used the coefficient of determination (R^2, a measure of random prediction error) and normalized mean absolute error bias (a measure of systematic prediction error) as performance metrics. The coefficient of determination and normalized absolute mean error (simply referred to as bias hereafter) are, respectively, defined in Eqs. 8 and 9:

$$R^2 = \frac{\sum(\hat{Y}-Y)^2}{\sum Y^2 - n\bar{Y}^2} \tag{8}$$

$$\text{Normalized mean absolute error} = \left|\left(\frac{\bar{\hat{Y}}-\bar{Y}}{\bar{Y}}\right)\right| \tag{9}$$

where $\hat{Y}$ and Y are the ANN-estimated and actual output, respectively, and $\bar{\hat{Y}}$ and $\bar{Y}$ are their respective means.

The results of the ANN training and testing for the four cases are shown in Tables 3A–D. The results for biased data (Case 1 through Case 3) (Tables 3A–D) show that the MSE (conventional ANN) performs very well in training, as shown by high R^2 (close to one) and low bias (close to zero). However, the results for the test phases deteriorate as test data are sampled differently than the training data from the sample space (total data set). Case 3 in general has relatively better performances for all parameters (in terms of higher R^2 and lower biases) compared with Cases 1 and 2. The R^2 value dipped mainly for Cases 1 and 2 with values as low as 0.628 (Case 1; $\log_{10}n$; Table 3iii). Performance in terms of the bias analogous to the R^2 performance: There are larger biases in the test phase for most of the cases (mostly greater than one order of magnitude) compared with the training phases (Tables 3A–D). The observed dip in performance statistics (R^2) in test phases of Cases 1 and 2 compared with Case 3 is consistent with the differences in the mean of the parameters (output data) in training and test phases in these three cases (Table 2).

The REF (regularized ANN), which penalizes difference in target means in the test phase, improves the statistics of the test phase as compared with the MSE (conventional ANN), as shown in Tables 3A–D. Although improvement in R^2 value could be seen in all biased cases (with the exception of Case 3; Θ_r), the improvement is apparent mainly for the first two cases with up to 62.5% improvement in R^2 value (Case 2; $\log_{10}\alpha$) over training phase. Similarly, the performance of regularized ANN

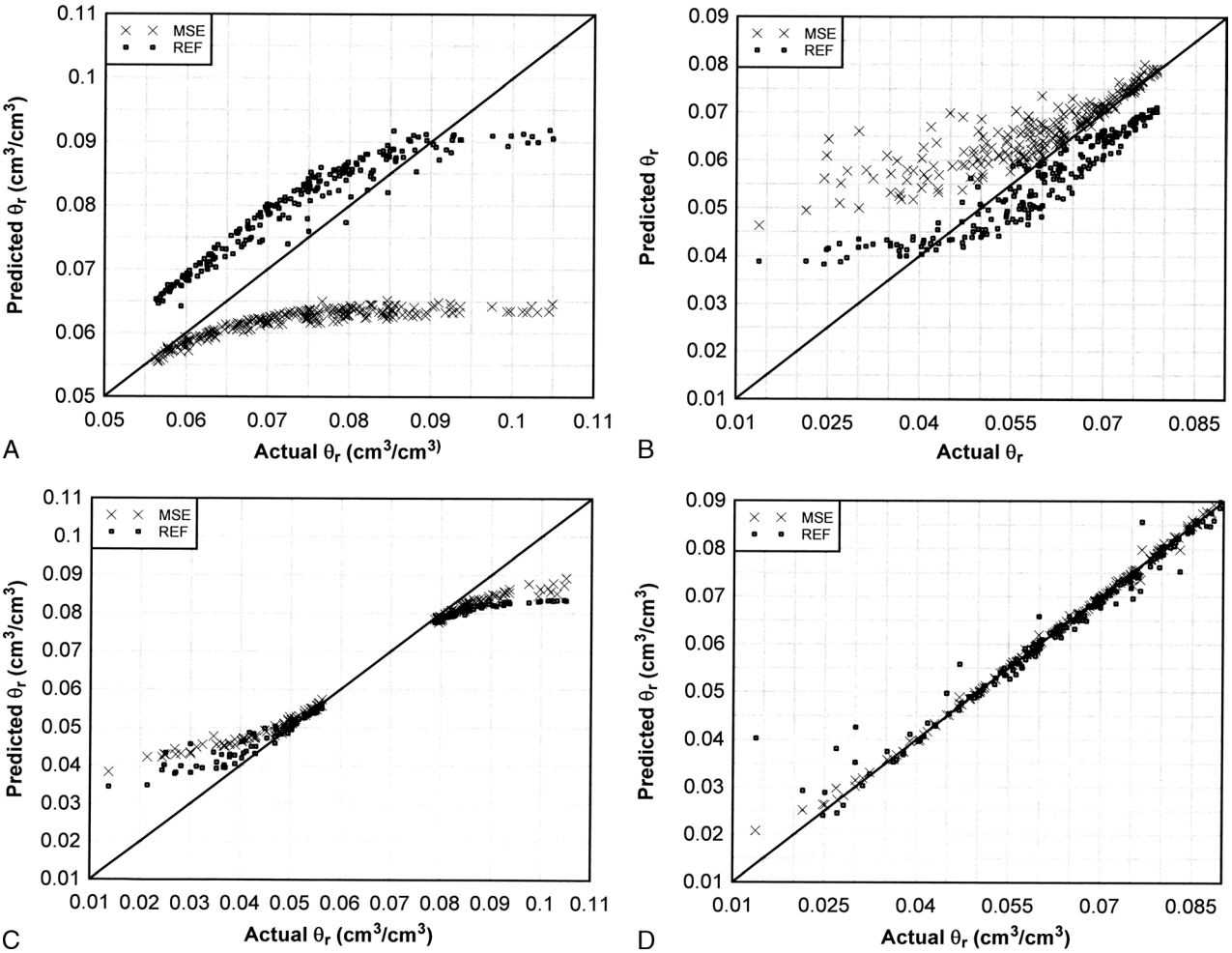

FIG. 1. Comparison of actual and predicted residual water content (Θ_r) in test phase (A) Case 1, (B) Case 2, (C) Case 3, and (D) Case-4.

is better through reduction of bias by 40% to 85% compared with the conventional ANN. Obviously, as the training targets to retain statistical properties of the test phases through penalty terms, the effect of introducing a mean penalty term is to correct the bias in the test phase, while introducing bias to the training result. Thus, the bias in the training phase is much lower in the conventional ANN compared with the regularized one. The improvement in the bias in the test phase of the regularized ANN is largely found to be dependent on differences in parameter mean values between training and testing phases. Improvements are superior for cases where differences in mean between training and test phases are higher (e.g., Cases 1 and 2; see Table 3i–iv). For Case 3, such differences in mean are the lowest among the biased cases; thus, both ANN are satisfactory. However, the regularized ANN shows better results in test phase for $\log_{10}n$, $\log_{10}\alpha$, and $\log_{10}K_s$ in terms of both R^2 value and bias (Table 3).

Figures 1A, 2A, 3A, and 4A illustrate results for test phases of Case 1 for parameters Θ_r, $\log_{10}K_s$, $\log_{10}n$, and $\log_{10}\alpha$, respectively. Although results in test phases are not perfect for either MSE or REF and bias is present, the reduction of bias is apparent in REF results. For the higher values of test data (within the considered cases), the estimated values in the conventional ANN tend to reach saturation (i.e., estimated value stay to a maximum and does not increase with actual value; e.g., Fig. 1A). This is attributed to the inherent limitation of ANN to predict beyond what it has learned in training phase and is related to both data scaling (min-max) and the nature of the sigmoid function (Minns and Hall, 1996). Literature in flood forecasting (where forecasting beyond training phase is of particular interest) shows a few studies devoted to such issues. While having a clipped linear function in the output node is considered to have beneficial effects on this problem (Karunanithi et al., 1994), reduced scaling in the training phase data (i.e., scaling training data to lie in a reduced range of sigmoidal activation function [e.g., by 10%–20%] to account for potential change in the test phase) has been recommended by many researchers (e.g., Imrie et al., 2007). The results, however, have not been effective with the conventional ANN training (Imrie et al., 2000). Because in Case 1 (Figs. 1A, 2A, 3A, and 4A) training is with the first quartile data (lower data range), the conventional ANN is unable to match higher value in test phases and thus tends to underestimate. Also, saturation of estimated parameter values is seen for the upper range of actual parameter values within the same quartile. The regularized ANN, on the other hand, is seen to reduce bias and improve results with some overestimation in the lower and underestimation for the higher range of observed values. Although saturation in the conventional ANN also occurs with the regularized ANN (Figs. 1A, 2A, 3A, and 4A), bias correction relatively minimizes the effect. The saturation pattern

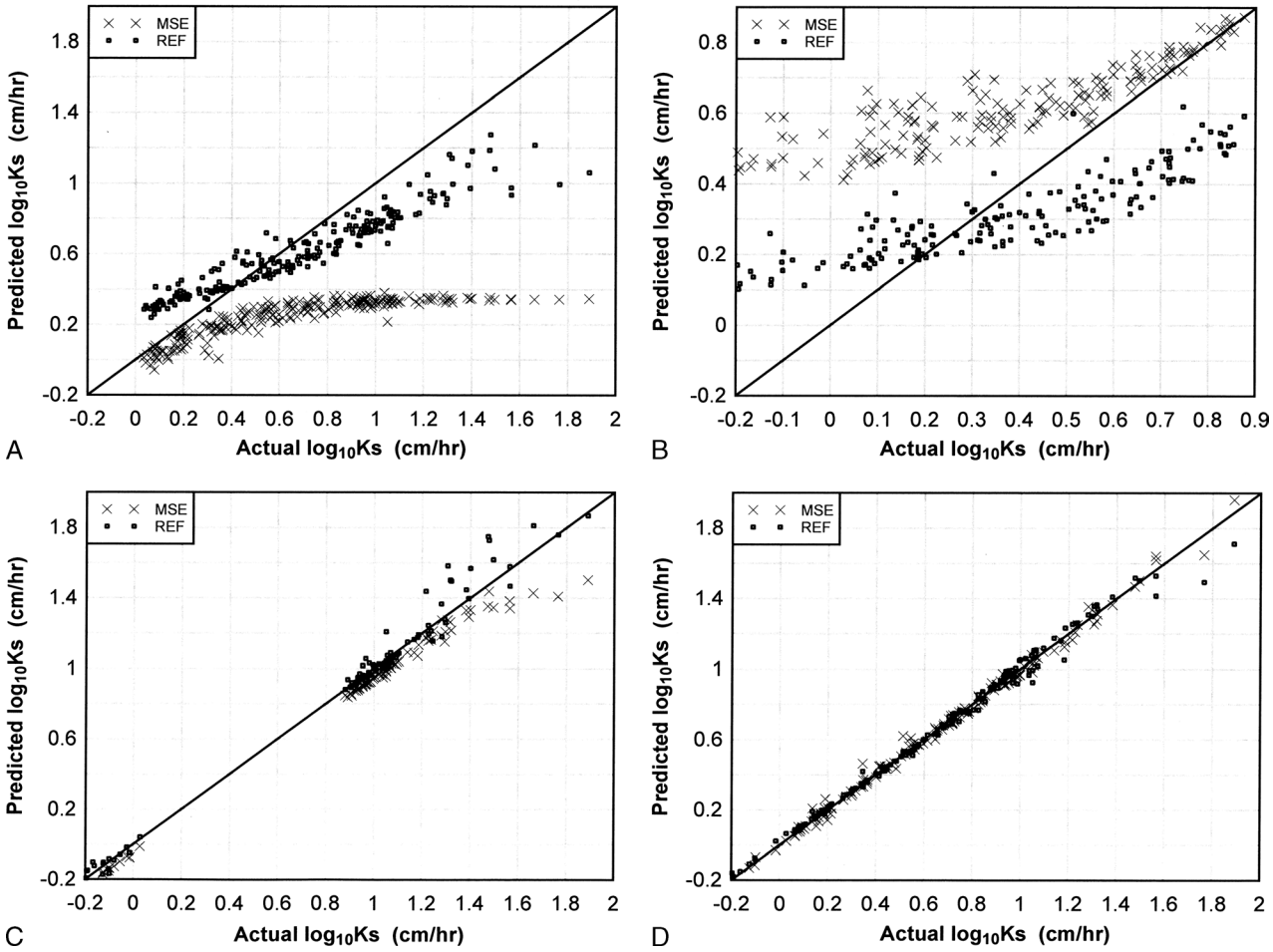

FIG. 2. Comparison of actual and predicted saturated hydraulic conductivity ($\log_{10}K_s$) in test phase (A) Case 1, (B) Case 2, (C) Case 3, and (D) Case-4.

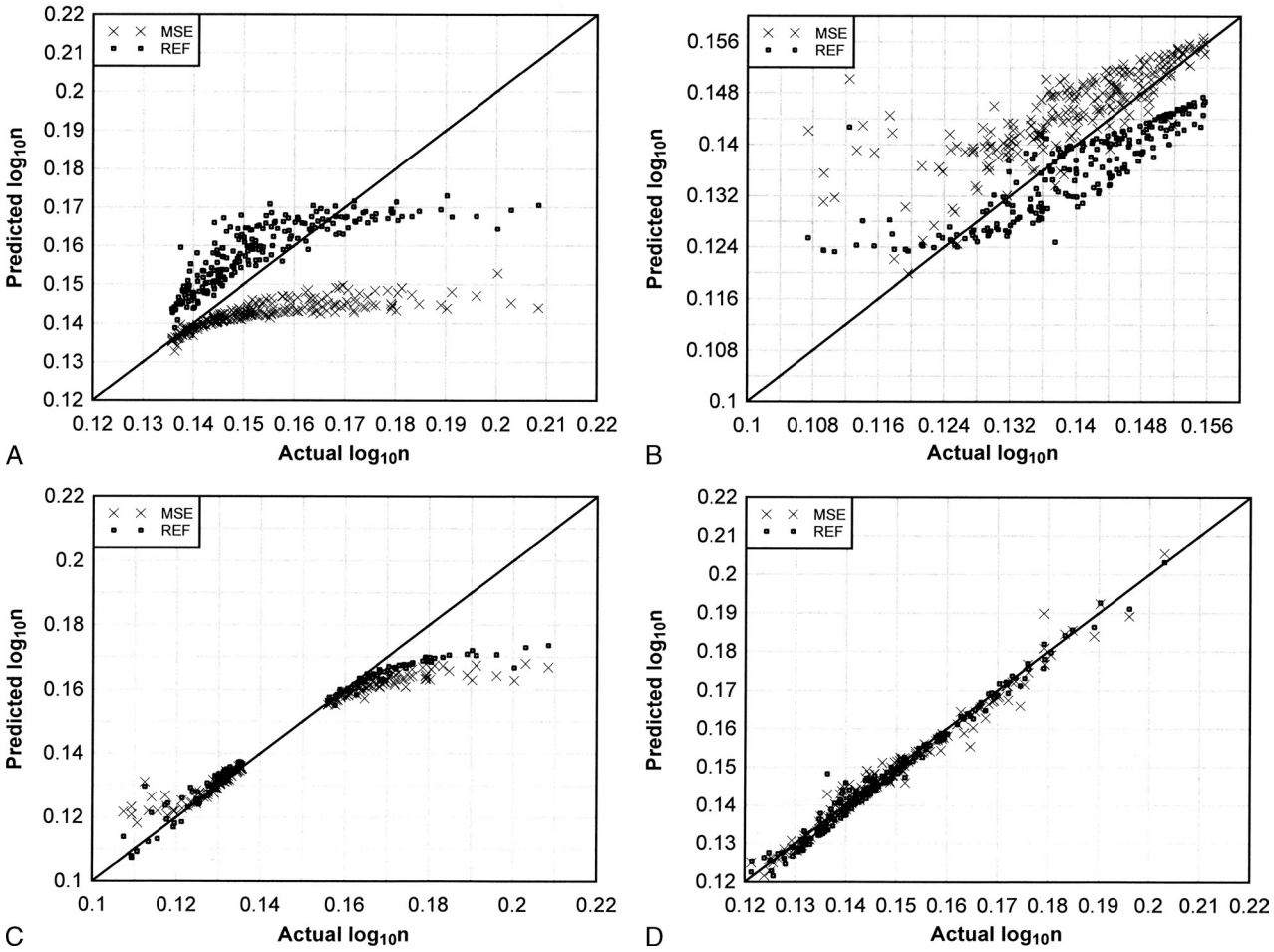

FIG. 3. Comparison of actual and predicted parameter $\log_{10}n$ in test phase (A) Case 1, (B) Case 2, (C) Case 3, and (D) Case-4.

is altered through upward shifting of estimated values to minimize bias (Figs. 3A and 4A) and occurs at relatively higher actual parameter values (Figs. 1A and 2A).

Similarly, Case 2 (Figs. 1B, 2B, 3B, and 4B) shows a similar result pattern to Case 1. However, saturation in the ANN result occurs in the lower actual values, reflecting the limit in the lower range of data in the training phase (training was with last quartile data). Although both ANN overestimate true values for lower actual parameter values, the regularized ANN does less so. For the upper range of actual values of the parameters, the regularized ANN underestimates, whereas the conventional one overestimates. Overall, performance of the regularized ANN is better as the improvement in the mean error with regularized ANN is apparent (Figs. 1B, 2B, 3B, and 4B; Table 3).

Case 3, which represents training in the interquartile range and testing with the first and last quartile data, deviates from the actual values for both high and low range of actual values of the parameters (Figs. 1C, 2C, 3C, and 4C). As in the former cases, the improvement in the mean error is also apparent in these figures, particularly for $\log_{10}n$, $\log_{10}K_s$, and $\log_{10}\alpha$.

For randomly sampled data (Case 4), however, the results are good in both test and training phases for both conventional and regularized ANN (Table 3, Case 4) as reflected by very high R^2 ($\geq$0.96 in both training and test phases) and almost similar low biases. Results for parameters for Case 4 are shown in Figs. 1D, 2D, 3D, and 4D, which show that both conventional and regularized ANN estimates followed the 45-degree line and thus worked equally well. Again, as can be seen from these figures, minor deviations from the target mean were observed, which were limited mostly to the extreme (low or high) range of parameters in the considered cases (Figs. 1D, 2D, and 4D). In the random sampling mode, data are picked randomly, so the mean of the sampled data for training is not much different from the rest (test phase). Thus, conventional MSE works equally well, and no additional advantage is apparent from REF.

One caveat of the example demonstrated here is that we have selected two extremely difficult cases where the training data are very different from test data (Cases 1 and 2). That poses some additional complications worth mentioning here. As training progresses, the training algorithm forces the predicted and actual output data close together. However, this introduces a large bias in the target mean in the test phase, thus forcing the training to readjust to new weights that can better take care of the mean in the test phase data. Thus, as no further improvement is achieved in the cross-validation data, the training process is terminated. This introduces bias in the training data set while correcting at the test or implementation phase (Table 3). This is not true for the less stringent case (Case 3), where the bias in the training phase with REF is also low (Table 3).

CONCLUSIONS

In this study, we introduced a new ANN training algorithm with REF to enable the use of available information on output

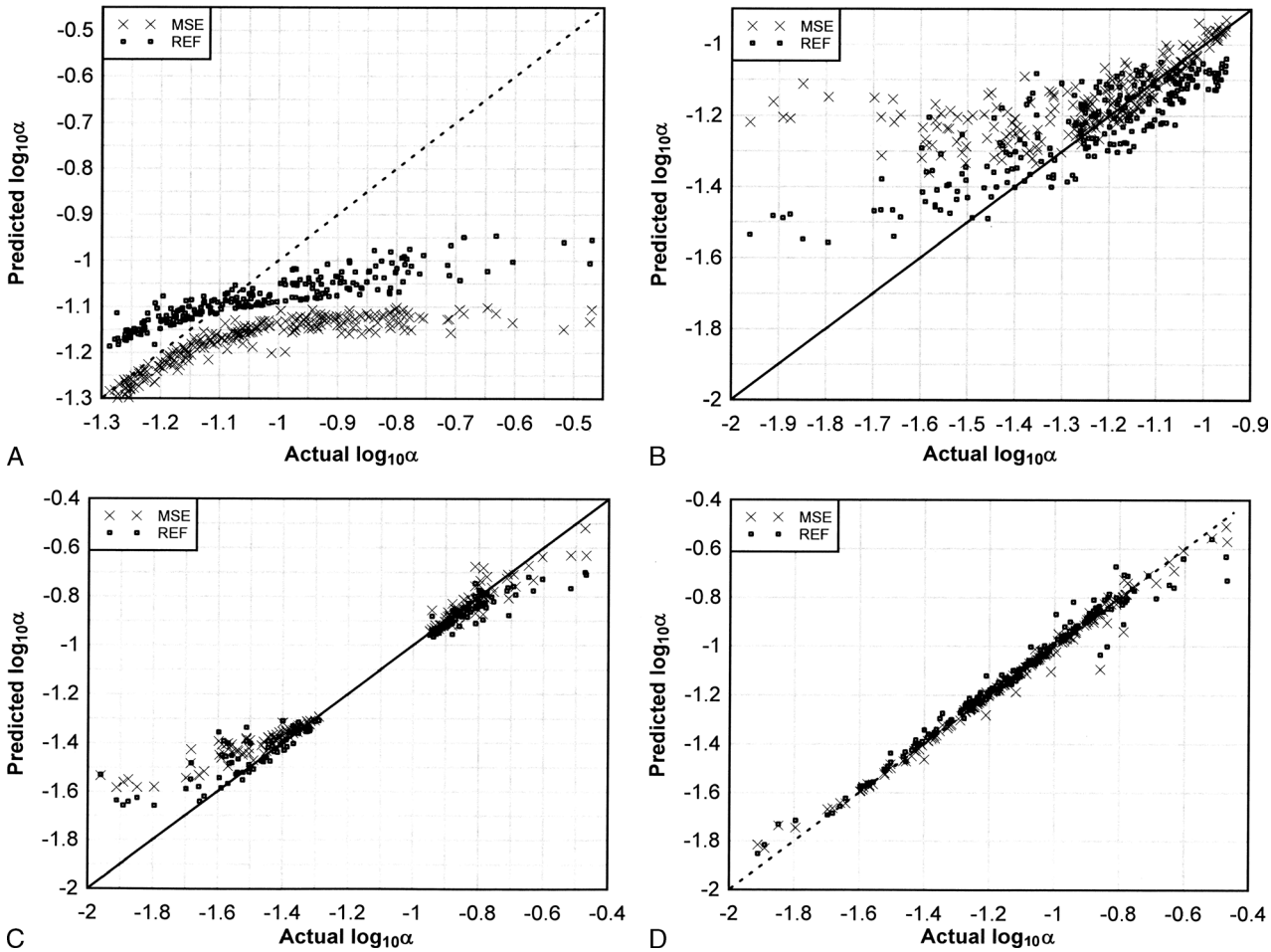

FIG. 4. Comparison of actual and predicted parameter $\log_{10}\alpha$ in test phase (A) Case 1, (B) Case 2, (C) Case 3, and (D) Case-4.

variables (e.g., mean). This approach is aimed at directly correcting bias in the test phase. Based on our results of four cases (three biased and one unbiased) of ANN experiments with four synthetic SHP data sets, we found the proposed algorithm to be effective. Specifically, the proposed regularized ANN was found highly effective in minimizing bias compared with the conventional ANN, particularly for highly biased data. The regularized ANN showed 40% to 85% less bias in highly biased data such as training with data from the first and last quartiles and testing on, respectively, last and first quartiles, where the difference in the target mean in the training and test phases is very high. Although saturation of the estimated values in test phase for higher and/or lower range of actual data was seen for both types of ANN, bias correction had the effect of relatively minimizing in results of regularized ANN. The new ANN training algorithm was found to perform better for those cases with relatively larger difference in target means of training and test data sets. However, training with both the regularized and conventional ANN resulted into equally good results in test phases when ANN models were trained with unbiased (statistically inferred from Wilcox rank sum test with <4.5% normalized mean differences in the training and test phases in the parameter values) data obtained from random sampling.

The developed algorithm has shown potential for improved estimation and especially for bias correction in the estimated values when there are significant differences in the mean in the training and test phases for a particular soil type and is thus useful to address problems such as scale-induced bias within the same soil group, as in Jana et al. (2007). For heterogeneous soil (different soil groups) with inherently different input-SHP physical relationships, it can be expected that any improvement in the mean error might be shadowed by the lower prediction accuracy (lower R^2 value). However, although not presented here, there is also potential to introduce penalty terms for higher-order moments in the proposed training algorithm. This should address some of the limitations arising from the correction on the target mean (i.e., first-order moment) only.

ACKNOWLEDGMENTS

The authors are grateful to Philip Meyer at the Pacific Northwest National Laboratory for his support to make this study possible. The authors also thank two anonymous reviewers for their constructive comments and suggestions.

REFERENCES

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, New York.

Børgesen, C. D., and M. G. Schaap. 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. Geoderma. 127:154–167.

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

Demuth, H., and M. Beale. 1992. *Neural Network Toolbox Manual*. MathWorks Inc., Natick, MA.

D'Agostino, R. B., A. Belanger, and R. B. D'Agostino. 1990. A suggestion for using powerful and informative tests of normality. Am Stat. 44:316–321.

Hornik, K., M. Stinchocombe, and H. Whilete. 1989. Multilayer feedforward networks are universal approximators. Neural Networks. 2:359–366.

Imrie, C. E., S. Durucan, and A. Korre. 2000. River flow prediction using artificial neural networks: Generalisation beyond the calibration range. J. Hydrol. 233:138–153.

Ines, A. V. M., and J. W. Hansen. 2006. Bias correction of daily GCM rainfall for crop simulation studies. Agr. Forest Meteorol. 138:44–53.

Jana, R. B., B. P. Mohanty, and E. P. Springer. 2007. Multiscale pedotransfer functions for soil water retention. Vadose Zone J. 6(4):868–878.

Jana, R. B., B. P. Mohanty, and E. P. Springer. 2008. Multiscale Bayesian neural networks for soil water content estimation. Water Resour Res. 44(8):W08408.

Karunanithi, N., D. Whitley, and K. Bovee. 1994. Neural networks for river flow prediction. J. Comput. Civil Eng. 8:201–220.

Koekkoek, E. and Bootlink, H. 1999. Neural network models to predict soil water retention. Eur. J. Soil Sci. 50:489–495.

Maier, H. M., and G. C. Dandy. 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environ. Model. Softw. 15(1):101–124.

Masters, T. 1993. Practical Neural Network Recipes in C^{++}. Academic Press Inc.

McKay M. D., W. J. Conover, and R. J. Beckman. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2):239–245.

Merdun, H., Ö. ÇInar, R. Meral, and M. Apan. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil and Tillage Res. 90:108–116.

Meyer, P. D., M. L. Rockhold, and G. W. Gee. 1997. Uncertainty analysis of infiltration and subsurface flow and transport for SDMP sites, NUREG/CR-6565, PNNL-11705. U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, DC.

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

Minasny, B. and A. B. McBratney. 2002. The neuro-M method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66:352–361.

Minasny, B. 2007. Predicting soil properties. Jurnal Ilmu Tanah Lingkungan. 7(1):54–67.

Minns, A. W., and M. J. Hall. 1996. Artificial neural networks as rainfall-runoff models. Hydrol. Sci. J. 41(3):399–417.

Mualem, Y. 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12(3):513–522.

Nemes, A., M. G. Schaap, and J. H. M. Wosten. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67:1093.

Pachepsky, Y., M. G. Schaap. 2004. Data mining and exploration techniques. *In* Y. Pachepsky, and W. J. Rawls (eds.). Developments in Soil Science. Development of Pedotransfer Functions in Soil Hydrology. Elsevier, San Diego. pp. 21–32.

Pachepsky, Y. A., D. Timlin, and G. Várallyay. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60:727–773.

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

Saxton K. E., W. J. Rawls, J. S. Romberger, and R. I. Timlin. 1986. Estimating generalized soil water characteristics from texture. Soil Sci. Soc. Am. J. 50:1031–1036.

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. and F. J. Leij. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163:765–779.

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

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. 251:163–176.

Stumpp, C., S. Engelhardt, M. Hofmann, and B. Huwe. 2009. Evaluation of pedotransfer functions for estimating soil hydraulic properties of prevalent soils in a catchment of the Bavarian Alps. Eur. J. Forest Res. 128: 609–620.

Tamari, S., J. H. M. Wösten, and J. C. Ruiz-Suarez. 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. Soil Sci. Soc. Am. J. 60:1732–1741.

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

van Genuchten, M. Th., and F. J. Leij. 1992. On estimating the hydraulic properties of unsaturated soils. *In* van Genuchten, M. Th., Leij, F. J., and Lund, L. J. (eds) Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA, 1989:1–14, 11–13.

Vereecken, H., M. Weynants, M. Javaux, Y. Pachepsky, M. G. Schaap, and M. Th. Genuchten. 2010. Using pedotransfer functions to estimate the van Genuchten-Mualem soil hydraulic properties: A review. Vadose Zone J. 9:795–820.

Woodbury, A. D. 1989. Bayesian updating revisited. Math. Geol. 21(3): 285–308.

Wösten J. H. M., YaA . Pachepsky, and W. J. Rawls. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251:123–150.

Ye, M., R. Khaleel, M. Schaap, and J. Zhu. 2007. Simulation of field injection experiments in heterogeneous unsaturated media using cokriging and artificial neural network. Water Resour. Res. 43:W07413.