Vadose Zone Journal
ORIGINAL RESEARCH

Multimodel Simulation of Water Flow in a Field Soil Using Pedotransfer Functions

A. K. Guber,* Ya. A. Pachepsky, M. Th. van Genuchten, J. Simunek, D. Jacques, A. Nemes, T. J. Nicholson, and R. E. Cady

Calibration of variably saturated flow models with field monitoring data is complicated by the strongly nonlinear dependency of the unsaturated flow parameters on the water content. Combining predictions using various independent models, often called *multimodel prediction*, is becoming a popular modeling technique. The objective of this study was to compare different methods of multimodel simulation of the field soil water regime using pedotransfer functions (PTFs). We solved the Richards flow equation using HYDRUS-1D with parameter sets derived from 19 published PTFs and compared different methods of combining the simulation results from the 19 individual models by (i) using only the best model, (ii) using equal weights, (iii) regressing measured values to the results of the individual models, (iv) using singular-value decomposition (SVD) in the regression, (v) using Bayesian model averaging, and (vi) using weights derived from Akaike criteria. Data on soil water contents and basic soil properties at five depths along a 6-m transect in a layered loamy soil were used to calibrate the Richards equation and to develop the input for the PTFs. The SVD multimodel was the best method, with an accuracy of about 0.01 $m^3\ m^{-3}$ at the 35-cm depth and about 0.005 $m^3\ m^{-3}$ at greater depths for 30 d of monitoring and 13 mo of testing. This indicates that multimodeling in combination with monitoring of the soil water regime can be a viable approach to simulating water flow in the vadose zone.

ABBREVIATIONS: BMA, Bayesian model averaging; PTF, pedotransfer function; SVD, singular-value decomposition.

SOIL WATER flow simulations have a multitude of applications at various scales, from irrigation scheduling to obtaining global climate change estimates. Modeling of flow in variably saturated soils requires water retention and hydraulic conductivity parameters that are impractical to measure for large-scale projects. As opposed to saturated flow, the nonlinearity of the unsaturated flow constitutive (hydraulic) properties seriously complicates calibration of the variably saturated flow models against field monitoring data.

Pedotransfer functions are routinely used to relate the hydraulic parameters to readily available data on soil properties that can be found on soil maps or extracted from soil survey reports. Since pedotransfer functions are empirical regression-type relationships, their accuracy outside of their development region is essentially unknown. A wealth of pedotransfer information has recently accumulated in nearly all parts of the world. Unfortunately, no good method currently exists to decide which pedotransfer function model should be used for a specific site or application (Pachepsky and Rawls, 2004).

Climate predictions faced similar uncertainties in model selection in the 1980s. To deal with these uncertainties, multimodel prediction emerged as a popular technique in climate prediction (Barnston et al., 2003; Palmer et al., 2000, 2004; Shukla et al., 2000). The objective of multimodel prediction is to reduce modeling errors by combining forecasts of various independent models. Since its introduction by Bates and Granger (1969), multimodel prediction has been subject to much debate that can be summarized into questions: (i) is a multimodel prediction better than the single best forecast, and (ii) what is the best approach to weigh predictions obtained with the different models? Regarding the first question, several studies (Kharin and Zwiers, 2002; Krishnamurti et al., 2000; Palmer et al., 2004) have reported systematic improvements in the forecasts with multimodel ensemble prediction, while others have argued that the benefit of multimodel prediction in seasonal predictions is marginal compared with that of using the best single model (Doblas-Reyes et al., 2000; Graham et al., 2000; Peng et al., 2002). Regarding the second question, some studies have demonstrated that proper weight selection leads to relatively better forecasts compared with simple averaging of the predictions from

A.K. Guber and J. Simunek, Dep. of Environmental Sciences, Univ. of California, Riverside, CA 92521; A.K. Guber and Y.A. Pachepsky, USDA-ARS, Environmental Microbial Safety Lab., BARC East, 173 Powder Mill Rd., Beltsville, MD 20705; M.Th. van Genuchten, Dep. of Mechanical Engineering, COPPE/LTTC, Federal Univ. of Rio de Janeiro, UFRJ, Rio de Janeiro, RJ, 21945-970, Brazil; D. Jacques, Belgium Nuclear Research Centre, SCK-CEN, Mol, Belgium; A. Nemes, USDA-ARS, Crop Systems and Global Change Lab., Beltsville, MD; and T.J. Nicholson and R.E. Cady, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Radioactive Waste, Rockville, MD. Received 17 Aug. 2007. *Corresponding author (Andrey.Guber@ars.usda.gov).

Vadose Zone J. 8:1–10
doi:10.2136/vzj2007.0144

several models (Krishnamurti et al., 2000; Pavan and Doblas-Reyes, 2000; Yun et al., 2003), while others have shown that weighting the forecasts from individual models is not better than simple averaging (Kharin and Zwiers, 2002; Peng et al., 2002; Doblas-Reyes et al., 2005). The various studies suggest that the answer to these two questions may well depend on the number of data points available for hindcasts (i.e., for evaluation of the models with data from the past).

The efficiency of multimodel predictions was demonstrated for streamflow forecasts by Regonda et al. (2006). Multimodel prediction methods are now slowly also being used in groundwater modeling. For example, Ye et al. (2004) suggested using the weighted results of several spatial variability models for unsaturated fractured tuff to run flow simulations for situations where standard information criteria provide an ambiguous ranking of the models such that it does not justify selecting one of them and discarding all the others. Methods of weighting predictions obtained from different groundwater flow models have been discussed in detail by Poeter and Anderson (2005). Recently, the feasibility of applying multimodel simulations to water flow in the vadose zone was demonstrated (Guber et al., 2006).

The objective of this study was to compare different methods of multimodel simulation of the field soil water regime using PTFs. We applied multimodeling using a large database of information on unsaturated soil water dynamics along a 6-m transect.

Methods for Multimodel Simulations

We consider the case where several PTF models exist to estimate parameters in the constitutive relationships (soil hydraulic properties) needed for application of the governing Richards equation to particular variably saturated flow problems. The multimodel simulation may then mean either (i) combining the output of several PTFs for the hydraulic parameters into a single parameter set and then running the flow model with this parameter set, or (ii) running the flow model with outputs of individual PTFs for the hydraulic parameters and then combining the obtained outputs of the flow model. In this study, we considered multimodel simulations in the sense of the second option, which is commonly used in meteorological predictions. Combining simulation results from N models is performed as

$$S = \overline{Y} + \sum_{i=1}^{N} w_i \left(F_i - \overline{F}_i \right) \quad [1]$$

where S is the multimodel simulation result, $\overline{Y}$ is the mean of the observed values during the training period, i is the model ($i = 1, 2, \ldots, N$), N is the total number of PTF models, F_i is the simulation result from model i, $\overline{F}_i$ is the mean of the simulation results obtained with model i during the training period, and w_i are the weights on the forecasts of individual models. Equation [1] shows that during the training period, the multimodel prediction equation relates deviations from the average of the observed values ($S - \overline{Y}$) with deviations from the average of the simulated values ($F - \overline{F}_i$).

Existing methods to combine predictions according to Eq. [1] differ in the way in which the weights w_i are obtained. Such methods have been reviewed by Clemen (1989), Burnham and Anderson (2002), Armstrong (2001), and Jolliffe and Stephenson (2003), among others. The review below briefly outlines the conceptual foundations of some of those methods and includes the approaches selected for comparison in our study.

Regression-Based Methods

If the measured values of the variable to be simulated are Y, then the weights w_i are determined by treating Eq. [1] as a multiple linear regression equation with $S - \overline{Y}$ as the dependent variable and $F_i - \overline{F}_i$ as the independent variable, and fitting the linear combination of multimodel predictions $\Sigma_{i=1}^{N} w_i(F_i - \overline{F}_i)$ to the observations $Y - \overline{Y}$. Such a method of obtaining multimodel forecasts has been called *superensemble forecasting* (Krishnamurti et al., 2000).

The main problem of using regression to search for w_i is the existence of relatively high correlations between F_i values. Having correlated independent variables in regression, or multicollinearity, does not preclude using the resultant regression for predictions within the range of observation (Neter and Wasserman, 1974); however, it leads to very inaccurate regression coefficients that are not easily interpreted and a matrix for the system of equations for calculating weights that is numerically close to singular (Kharin and Zwiers, 2002). Yun et al. (2003) showed that improvements in superensemble forecasts can be achieved by applying an SVD technique to solve the system of equations for the coefficients w_i.

Bayesian Model Averaging

Following Raftery et al. (2003), a basic premise is the assumption that for any given forecast, there is a "best" model; while we do not know what that model is, uncertainty about the best model can be quantified using Bayesian model averaging. Let the bias-corrected value to simulate be $Y - \overline{Y}$, and the bias-corrected simulations $f_i = F_i - \overline{F}_i$. Simulation f_i is then associated with a probability density function (PDF), $g_i = (Y - \overline{Y} \,|f_i)$, which is interpreted as the PDF of $Y - \overline{Y}$ conditioned on f_i given that f_i is the best simulation. The Baysian model averaging (BMA) predictive model is

$$p\left(Y - \overline{Y} \mid f_1, f_2, \ldots, f_N\right) = \sum_{i=1}^{N} w_i g_i \left(Y - \overline{Y} \mid f_i\right) \quad [2]$$

where w_i is the posterior probability of the simulation f_i being the best one and is based on the performance of model i in the training period. The w_i are probabilities that add up to one.

Raftery et al. (2003) indicated that it is often reasonable to use the normal distribution $N(f_i, \sigma_i^2)$ to approximate the conditional PDF $g_i = (Y - \overline{Y} \,|f_i)$. In that case, the BMA predictive mean is the conditional expectation of Y given the simulations, namely

$$\mathrm{E}\left[Y - \overline{Y} \mid f_1, f_2, \ldots, f_N\right] = \sum_{i=1}^{N} w_i f_i \quad [3]$$

The weights w_i and the variances σ_i^2 are found using the maximum likelihood method. The expectation maximization, or EM algorithm (McLachlan and Krishnan, 1997) can be used for the normal conditional PDFs (Raftery et al., 2003). The EM algorithm is a method for finding the maximum likelihood estimator when the problem can be recast in terms of "missing data." This algorithm is iterative, and alternates between two steps, the E (or expectation) step, and the M (or maximization) step. Assuming that the conditional PDF $g_i = (Y - \overline{Y} \,|f_i)$ is computed as

$$g_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{\left(f_i - Y + \overline{Y}\right)^2}{2\sigma_i^2} \right] \quad [4]$$

the algorithm starts with initial guesses for the weights w_i and variances σ_i. At the E step, estimates of the missing values $z_{ji}^{(m)}$ are computed as

$$z_{ji}^{(m)} = \frac{g\left(Y_j - \overline{Y} \middle| f_{ji}, \sigma_i^{m-1}\right)}{\sum_{i=1}^{N} g\left(Y_i - \overline{Y} \middle| f_{ji}, \sigma_i^{m-1}\right)} \quad [5]$$

where j is the number of the observation within the training period (j = 1, 2, …, K), i is the number of the model, and m is the number of the iteration. The M step then consists of estimating weights and variances as

$$w_i^{(m)} = \frac{1}{K}\sum_{j=1}^{K} z_{ji}^{(m)} \quad [6]$$

$$\sigma_i^{(m)} = \frac{\sum_{j=1}^{K} z_{ji}^{(m)} \left(Y_j - \overline{Y} - f_{ji}\right)^2}{\sum_{j=1}^{K} z_{ji}^{(m)}} \quad [7]$$

In general, Bayesian model averaging is designed and has been used to generate probabilistic forecasts. For that reason, this approach appears to have wider applicability in probabilistic forecasting than deterministic regression-based averaging (Rajagopalan et al., 2002).

Using Information Theory

Following Burnham and Anderson (2002), Poeter and Anderson (2005) suggested using the Akaike information criterion to derive the weights for individual models as

$$w_i = \frac{\exp\left(-0.5\Delta_i\right)}{\sum_{l=1}^{N} \exp\left(-0.5\Delta_l\right)} \quad [8]$$

where

$$\Delta_i = \mathrm{AICc}_i - \mathrm{AICc}_{\min} \quad [9]$$

AICc_i is the value of the Akaike criterion for the ith model, and $\mathrm{AICc}_{\min}$ is the minimum AICc value of all models in the set. The AICc value for a model is given by

$$\mathrm{AICc} = n\log\left(\sigma^2\right) + 2k + \frac{2k(k+1)}{n-k-1} \quad [10]$$

where n is the number of observations, σ^2 is the residual variance estimated as the sum of squared residuals divided by n, and k is the number of estimated parameters. This methodology is efficient if models require estimating their parameters before the prediction averaging.

Multimodel Simulations

The following multimodel prediction methods were compared in our study: (i) using only the best model, (ii) assigning equal weights to all models, (iii) using the superensemble, (iv) using the superensemble with SVD to find weights, (v) using Bayesian model averaging, and (vi) using information theory. Initial estimates of the BMA weights were taken proportional to the inverse RMSEs of individual models. Singular-value decomposition was performed using routines from Press et al. (1992).

Experimental Data

The experimental field was located at Bekkevoort, Belgium, in a meadow at the bottom of a 4% slope. The soil was classified as a Eutric Regosol (FAO, 1975) or a Udifluvent (Soil Survey Staff, 1999). The top 1 m included three soil horizons: an Ap horizon between 0 and 25 cm, a C1 horizon between 25 and 55 cm, and a C2 horizon between 55 and 100 cm. A trench, 1.2 m deep and 8 m long, was excavated at the field site. Soil texture was measured with the pipette method following pretreatment with sodium hexametaphosphate. Textural classes were loam at the 15-, 35-, and 55-cm sampling depths and silty loam at 75- and 95-cm depths (Table 1). The grass cover was removed from the experimental area. A plastic sheet covered the side of the trench along which instrumentation was installed. Volumetric water contents were measured with time domain reflectometry (TDR). Sixty TDR probes (two rods 25 cm long, 0.5-cm rod diameter, 2.5-cm rod spacing) were installed along the trench at 12 locations spaced 50 cm apart laterally, at five depths of 15, 35, 55, 75, and 95 cm. The TDR measurements were performed with a Tektronix (Beaverton, OR) 1502B cable tester. The automated system of Heimovaara and Bouten (1990) was used to control, retrieve, store, and analyze measurements of the travel time of the electromagnetic waves along the TDR rods. The time difference between two measurements for the same probe was 2 h. Rainfall was measured and recorded continuously near the trench across a catch area of 200 cm^2 (Fig. 1).

The trench was filled after all devices were installed. A thin layer of gravel (1–2 cm) was evenly distributed over the study area to (i) decrease the erosive effect of rain impact on the bare soil surface, (ii) minimize evaporation from the soil surface, and (iii) decrease the growth of weeds on the experimental plot. Weeds were regularly removed from the site during the summer. Field measurements started on 11 Mar. 1998 (Day 0). We used 250 d of data. Additional details about the site and the experiment are given by Jacques (2000) and Jacques et al. (2001).

Pedotransfer Functions and Flow Simulations

The literature was searched for pedotransfer functions to estimate the soil water retention and hydraulic conductivity properties from soils data available at the site. We used only PTFs that had been developed from relatively large (>200 samples) databases. To estimate soil water retention, we selected 19 PTFs developed in different regions. Equations for the PTFs were taken from the appendices of Guber et al. (2006) and Pachepsky et al. (2006). Five of the PTFs (those by Campbell and Shiozawa [1992], Mayr and Jarvis [1999], Rawls and Brakensiek [1985],

TABLE 1. Average values of soil properties at the monitoring depths.

Depth	No. of samples	Soil textural fraction					Bulk density	Organic C
		>50 µm	50–20 µm	20–10 µm	10–2 µm	<2 µm		
cm		% (w/w)					g cm^{-3}	% (w/w)
15	7	58.6	19.3	6.4	4.5	11.1	1.42	2.2
35	8	56.7	18.9	7.8	3.2	13.3	1.54	0.8
55	5	57.3	17.6	6.6	3.7	14.8	1.53	0.4
75	3	49.6	21.2	7.9	4.4	17.4	1.53	0.3
95	4	43.8	30.3	7.4	4.5	14.0	1.53	0.6

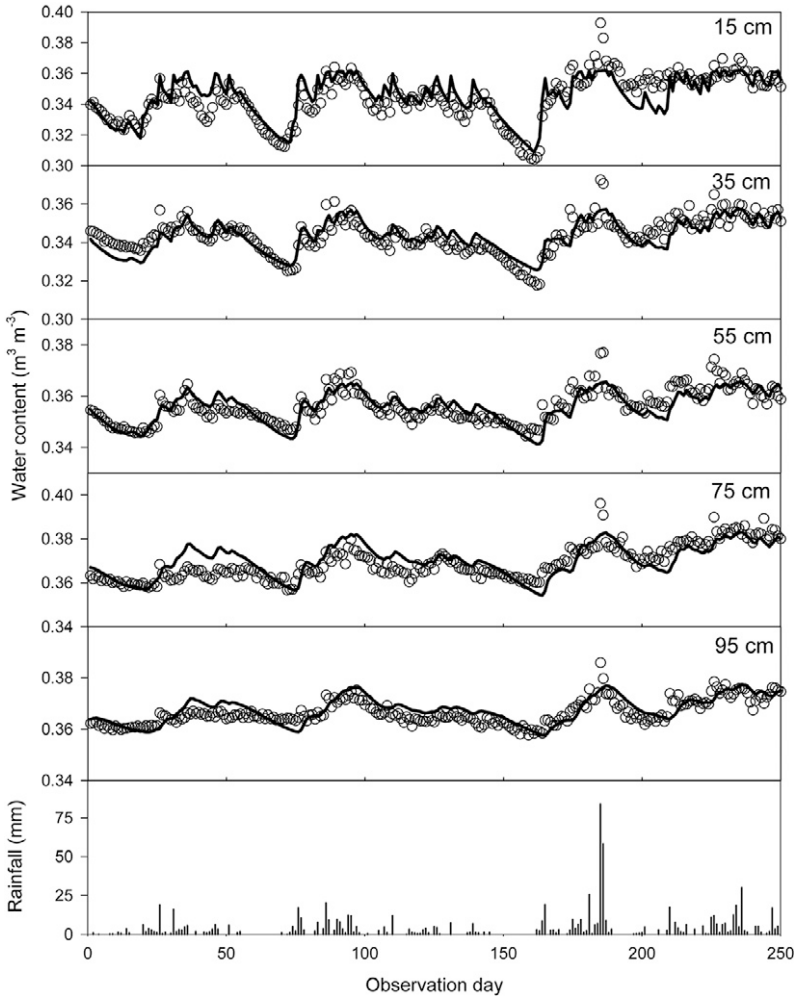

FIG. 1. Observed (symbols) and simulated (using the calibrated model, lines) daily average water contents. Observed values are averages across the transect for each depth and each day.

TABLE 2. Enumeration of pedotransfer functions (PTFs) investigated in this work.

No.	PTF	PTF predictors					
		Clay	Silt	Sand	Bulk density	Organic matter	Itop†
1	Rawls et al. (1982)	+	+	+	+		
2	Saxton et al. (1986)	+		+			
3	Williams et al. (1992)	+		+		+	
4	Campbell and Shiozawa (1992)	+		+	+		
5	Rawls et al. (1983)	+	+	+	+	+	
6	Verekeen et al. (1989)	+		+	+	+	
7	Gupta and Larson (1979)	+	+	+	+	+	
8	Baumer (1992)	+		+	+	+	
9	Varallyay et al. (1982)	+		+	+		
10	Canarache (1993)	+			+		
11	Peterson et al. (1968)	+					
12	Rajkai and Várallyay (1992)	+		+	+	+	
13	Tomasella and Hodnett (1998)	+	+				
15	Rawls and Brakensiek (1985)	+		+	+		
16	Hall et al. (1977)	+	+	+	+		
17	Williams et al. (1992)	+		+	+		
18	Wösten et al. (1999)	+	+	+	+	+	+
18	Mayr and Jarvis (1999)	+	+	+	+	+	
19	Rosetta (Schaap et al., 2001)	+	+	+	+		

† Variable equal to zero for subsoil and one for topsoil.

Saxton et al. [1986], and Williams et al. [1992]) estimated the parameters of the equation of Brooks and Corey (1964):

$$\frac{\theta-\theta_r}{\phi-\theta_r}=\begin{cases}\left(\frac{h_b}{h}\right)^{\lambda}, & h>h_b\\ 1, & h\le h_b\end{cases} \quad [11]$$

while four other equations (Varallyay et al., 1982; Vereecken et al., 1989; Wösten et al., 1999) estimated the parameters of the van Genuchten (1980) equation

$$\frac{\theta-\theta_r}{\theta_s-\theta_r}=\frac{1}{\left[1+(\alpha h)^n\right]^m} \quad [12]$$

where θ is the volumetric water content, h is the pressure head (taken here positive for unsaturated conditions), ϕ is the porosity, θ_r is the residual water content, h_b is bubbling pressure, λ is the pore size distribution index, θ_s is the saturated water content, and α, m, and n are empirical shape-defining parameters. Nine pedotransfer functions (Baumer, 1992; Canarache, 1993; Gupta and Larson, 1979; Hall et al., 1977; Petersen et al., 1968; Rajkai and Várallyay, 1992; Rawls et al., 1982, 1983; Tomasella and Hodnett, 1998) estimated water contents at several fixed pressure heads. The van Genuchten parameters θ_s, θ_r, α, and n in these cases were evaluated by fitting Eq. [12] to the water retention points obtained from those PTFs. The value of parameter m was calculated as $m = 1 - 1/n$. The residual water content θ_r was set to 0.001 m^3 m^{-3} and the saturated water content θ_s to the porosity ϕ for PTFs that evaluated values of the water content at only two pressure heads (15,000 and 330 cm). Two PTFs (Vereecken et al., 1989; Varallyay et al., 1982) used van Genuchten parameters assuming $m = 1$. To use these PTFs for the flow simulations, water contents were calculated at the pressure heads used by those researchers to derive the PTFs, after which we estimated θ_s, θ_r, α, and n by fitting Eq. [12], with $m = 1 - 1/n$, to the calculated water retention data. We also used the Rosetta software (Schaap, 2004) to generate van Genuchten parameters from texture and bulk density. Pedotransfer functions were enumerated as shown in Table 2. A FORTRAN code to estimate water retention with the pedotransfer functions of this study, except Rosetta, is available on request from the corresponding author.

The saturated hydraulic conductivity, K_{sat}, has also been estimated from a PTF. The literature contains only three K_{sat} PTFs that were developed or tested using large databases (Rawls et al., 1998; Wösten et al., 1999; Schaap et al., 2001). We focused on the multimodeling with different water retention PTFs, and used only the PTF developed by Rawls et al. (1998).

The HYDRUS-1D software (Simunek et al., 1998) was used to run the simulations. This software gives options to run simulations either with the Brooks–Corey water retention Eq. [11] or the van Genuchten

TABLE 3. Calibrated van Genuchten (Eq. [12]) and van Genuchten–Mualem (Eq. [14]) parameters for the Richards flow equation, and associated RMSEs for different soil horizons.

Depth	θ_r	θ_s	α	n	K_{sat}	l	RMSE
cm	$m^3\ m^{-3}$		cm^{-1}		$cm\ d^{-1}$		$m^3\ m^{-3}$
0–25	0.005 ± 0.220	0.363 ± 0.001	0.00110 ± 0.0004	1.80 ± 0.19	0.055 ± 0.006	0.01 ± 0.83	0.0088
25–45	0.086 ± 1.209	0.362 ± 0.008	0.00184 ± 0.0040	1.27 ± 1.14	50 ± 409	0.50 ± 36.4	0.0055
45–65	0.154 ± 1.150	0.367 ± 0.004	0.00123 ± 0.0044	1.53 ± 1.35	0.207 ± 0.300	5.01 ± 54.4	0.0038
65–85	0.108 ± 1.443	0.388 ± 0.008	0.00133 ± 0.0042	1.49 ± 1.44	0.167 ± 0.329	0.5 ± 24.9	0.0053
85–105	0.010 ± 0.453	0.386 ± 0.004	0.00155 ± 0.0016	1.20 ± 0.20	4.16 ± 7.64	18.4 ± 31.5	0.0043

Eq. [12]. The particular equation was selected depending on the invoked PTF. The unsaturated hydraulic conductivity (K) function was calculated using

$$K(h) = K_{sat} S_e^{2/\lambda + l + 2} \quad [13]$$

for the model of Brooks and Corey (1964), and

$$K(h) = K_{sat} S_e^{l} \left[1 - \left(1 - S_e^{1/m} \right)^m \right] \quad [14]$$

for the van Genuchten–Mualem model (van Genuchten, 1980), where

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} \quad [15]$$

and l is a pore-connectivity parameter. A separate simulation run was performed with HYDRUS-1D for each PTF. Water retention was estimated by applying the PTFs to the soil properties of each individual layer at the Bekkevoort field site (0–25, 25–45, 45–65, 65–85, and 85–150 cm; see Table 1).

We evaluated the weighting methods in terms of their accuracy (i.e., errors in reproducing the training or hindcast data sets), and reliability (i.e., errors in reproducing the test data sets). The training (i.e., determining the weights) was done with daily water contents using moving windows that were from 30 to 150 d long. All data outside the windows were used to test the multimodel prediction. Average and standard deviations of RMSEs for each window size were obtained separately for the testing and training sets.

The Richards water flow model was calibrated using complete time series of measured water contents, measured precipitation data, and evaporation rates estimated from water budget computations (Pachepsky et al., 2006). The calibration was used to evaluate the applicability of the Richards equation to the observed flow processes, and to compare the accuracy of the calibrated model with those of the multimodel predictions. The inverse solution option of HYDDRUS-1D was used for this purpose. The soil profile was subdivided into five layers: 0 to 25, 25 to 45, 45 to 65, 65 to 85, and 85 to 105 cm. The four parameters θ_r, θ_s, α, and n in Eq. [12] and the two additional parameters K_{sat} and l in Eq. [14] were estimated for each layer by fitting the Richards flow model to daily water content values averaged across the trench at each depth. Initial estimates of the parameters were obtained by using data on soil texture (Table 1) and the Rosetta PTF included in the HYDRUS-1D software. A total of 30 parameters were estimated during model calibration.

Results

Observed and simulated soil water contents are shown in Fig. 1. The topsoil exhibited more weather-related variations in the water content than the subsoil. Still, the relatively dry periods between Days 55 and 76 and between Days 140 and 166 are reflected in lower soil water contents at depths down to 75 cm.

Calibrations with the Richards flow model for the five-layer profile resulted in relatively high accuracy (Fig. 1). The RMSE values decreased with depth (Table 3); calibrated parameters along with their standard errors and linear tolerance intervals are shown in Table 3. The parameter θ_r was found to have the lowest accuracy. This was not surprising since most of the observed water contents and pressure heads were in the relatively wet range (data not further shown). In contrast, values of θ_s were very reliable. The parameters α and n could be defined reliably only for the top layer. For all other layers, the tolerance intervals were very wide, although the estimates themselves were acceptable. As indicated by Hill (1998), this suggests that the information content of data for the deeper layers was insufficient for conclusive evaluation. Since the top layer experienced more drying (Fig. 1), the simulations were found to be more sensitive to the retention shape parameters α and n. By comparison, most or all of the data in the deeper layers remained close to saturation, leading to unreliable estimates for α and n. The same is true for the hydraulic conductivity parameters K_{sat} and l. Still, the satisfactory agreement between the observed and calculated data does indicate applicability of the Richards equation to the flow processes at this site.

The HYDRUS-1D model (Simunek et al., 1998) was used next to simulate flow at the field site using hydraulic parameters obtained with the PTFs. An example of the results of these simulations is shown in Fig. 2 for the 15-cm depth. The scatter in the simulated water contents was found to be substantial. In general, the simulated soil water contents were lower than the observations and also decreased somewhat faster during dry periods. Temporal variations in the soil water content were

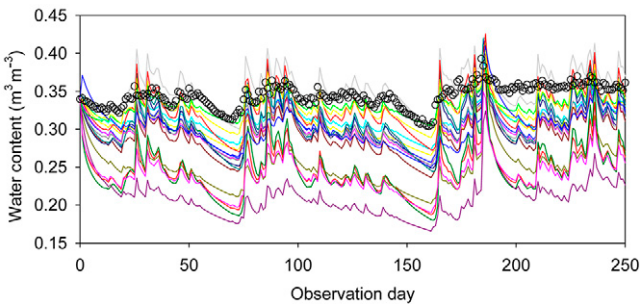

FIG. 2. Daily average water contents at the 15-cm depth; lines are simulations with 19 pedotransfer functions; circles are observed values.

more pronounced with some PTFs than with others; however, inspection of Fig. 2 shows that the different PTFs produced almost parallel time series that were shifted in various degrees with respect to each other. Computed correlation coefficients between the time series supported this finding. A color-coded matrix of correlation coefficients between the calculated water content time series using different pedotransfer functions is shown in Fig. 3 for the 15-cm depth. Correlation coefficients between simulated soil water contents exceeded 0.95 in >48% of the cases, while very few correlation coefficients were <0.5. Similar results were obtained for the other depths (data not shown).

The accuracy and reliability of different methods to build the multimodel prediction are compared in Fig. 4. The superensemble was found to be the worst weighting method, characterized by high accuracy and very low reliability (data not shown in Fig. 4 because they are outside the axis scales). Assigning equal weights (simple averaging) was the second worst method. This method also had testing RMSE values larger than the training RMSEs, although the difference was not as dramatic as with the superensemble method. The accuracy of the superensemble was the worst of all the methods. Bayesian model averaging fared only slightly better than simple averaging. Using only the best model was a reasonably good approach in terms of accuracy. The reliability of this method, however, was almost the same as with simple averaging and Bayesian averaging. The results of the information theory method were no different from using the best model. Using SVD in the superensemble appeared to be the best method in that both the accuracy and reliability RMSE values were almost two times less than those of all other methods considered in this study. An example in Fig. 5 of applying the SVD to the whole 250-d observation period as a training period illustrates the performance of this technique.

The results in Fig. 4 further show that training with 30 d of data leads to lower reliability than when 60 or 90 d of data are used in the training. We found a weak trend of increasing accuracy with the duration of training. Figure 4 also compares multimodel results with the results obtained with the Richards model calibrated on data from the entire observation period. The error bars for the calibrated model were obtained from the statistical distributions of RMSE for the same training and testing periods that were used for the multimodel predictions. The data in Fig. 4 show that the accuracy of the superensemble with SVD is comparable to the accuracy and reliability of the calibrated Richards equation, even when weights are derived from 30 d of observations. The overall accuracy and reliability of the SVD multimodel was slightly less, or comparable to, the accuracy of the calibrated model.

The reliability of the SVD method was found to be not random in time. The RMSE time series shown in Fig. 6 indicate that the SVD multimodel

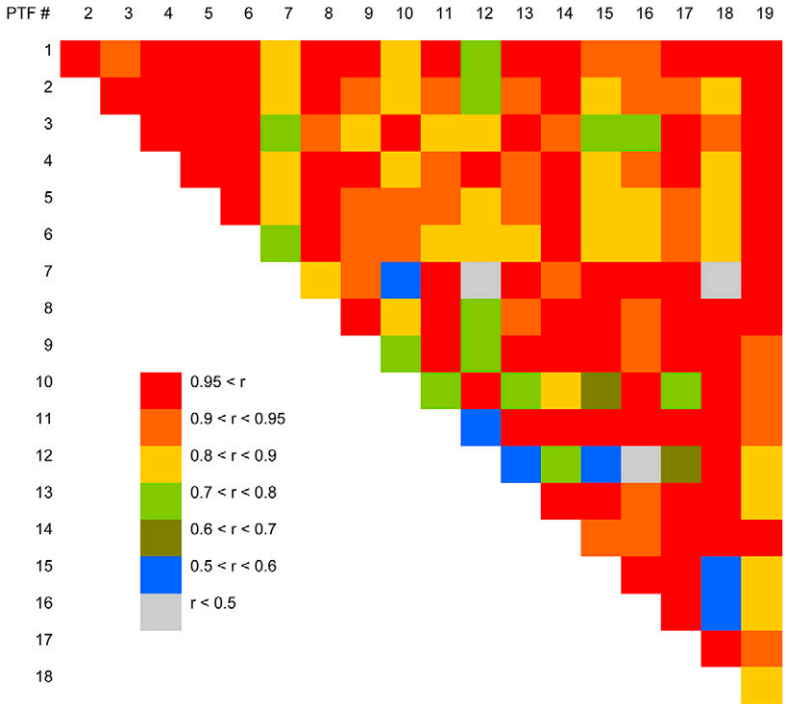

FIG. 3. Spearman correlations between time series of water contents at the 15-cm depth from simulations obtained with different pedotransfer functions (PTFs).

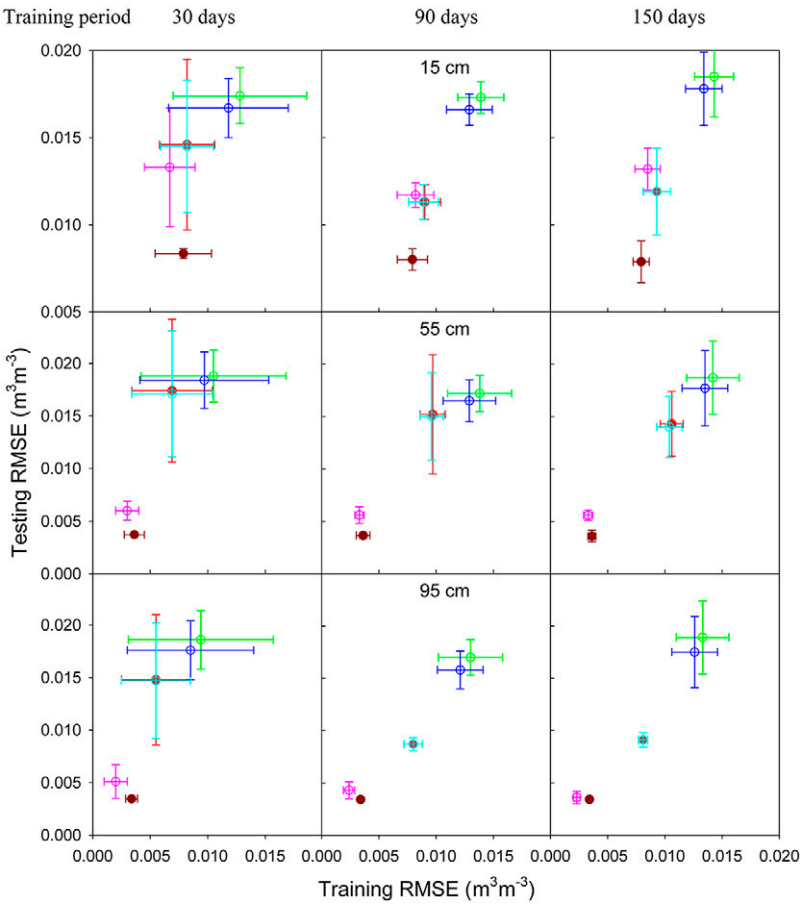

FIG. 4. Root mean squared errors of multimodel simulations of the training and testing data sets; results are for the best in training model (red), the arithmetic average of all individual models (green), the superensemble with singular-value decomposition (pink), Bayesian averaging (blue), the use of the information theory (cyan), and calibrated Richards equation (brown).

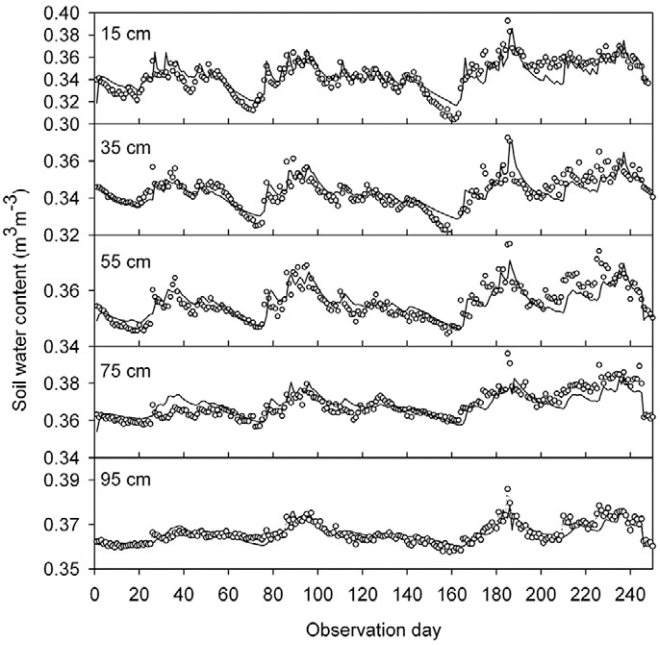

FIG. 5. Results of applying the superensemble with the singular-value decomposition to multimodeling of the whole observation period. Daily average water contents at 15-cm depth; lines are simulations and circles are observed values.

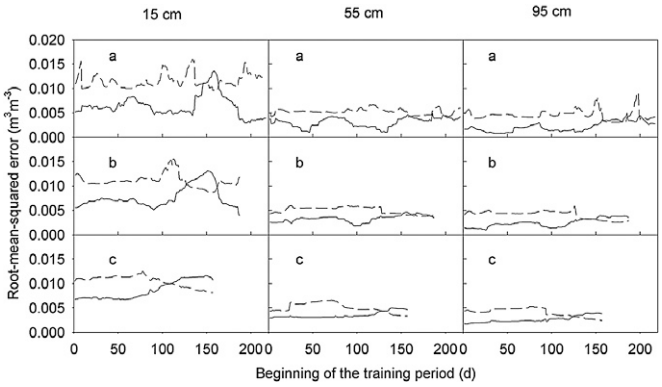

FIG. 6. Dependencies of root mean square errors in the training (solid line) and testing (dashed line) data sets on the beginning and duration of the training period for the superensemble with singular-value decomposition method for training periods of (a) 30, (b) 60, and (c) 90 d.

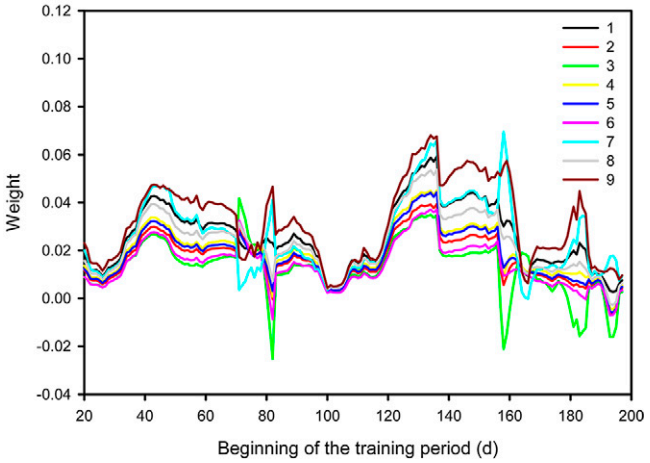

FIG. 7. Weights of the simulations with the first nine pedotransfer functions in the superensemble multimodel with singular-value decomposition obtained for the 15-cm depth with 30-d training periods.

was slightly less accurate, and slightly more reliable, in the fall toward the end of the experiments. This was observed for all depths (data partly shown in Fig. 6). Spearman correlation coefficients between SVD-simulated and measured soil water contents were between 0.6 and 0.8 at the 15-cm depth, between 0.5 and 0.8 at the 35-cm depth, and between 0.4 and 0.8 depending on the starting day of the training period.

Time series of weights of the individual models as a function of the beginning of the training period are shown in Fig. 7 for the first nine of 19 individual PTFs at the 15-cm depth. Inspection of the weights shows that none of the models was consistently ignored. Depending on the training period, the same PTF could have larger or smaller input into the multimodel prediction. While the weights of a few PTFs were sometimes negative, in most cases the PTF predictions were summed up with positive weights. Similar results were obtained for the other four depths (data not shown).

Discussion

Multimodel simulations rely on the assumption that the physics of each individual model is correct. Unfortunately, PTFs are merely estimators of model parameters, and as such do not affect the description of the physical processes involved, in this case the Richards equation based on mass conservation in combination with the Darcy–Buckingham law for the fluid flux. This flow model was applicable in principle since it could be fitted reasonably well to the data (Fig. 1). Our attempts to use multimodel simulations based on the Richards equation are hence warranted. The similarity in simulated physical processes caused strong correlations between predictions from individual models (Fig. 2 and 3), which is a typical feature of multimodel predictions.

Different methods used for the multimodel predictions were found to perform quite differently. The superensemble with SVD was found to be by far the best method in our study (Fig. 4). This conclusion cannot be generalized. We believe that a preliminary assessment of all methods should be performed for specific applications, at least until some experience is accumulated in using multimodels for simulations of flow in variably saturated soils. Also, the saturated hydraulic conductivity, K_{sat}, was estimated from a single PTF in this study, whereas two more PTFs (Wösten et al., 1999; Schaap et al., 2001) also could be used to expand the multimodel, which includes 57 rather than 19 individual models.

Values of weights in Eq. [1] were negative for some models when the regression-based methods were applied. Bayesian model averaging and Akaike criteria application do not allow negative weights. The regression-based methods treat weights as fitting coefficients that can have any sign. The reason for negative weights is that the model predicts behavior opposite to that measured. That may happen because of prediction errors of some individual PTFs. Consider a PTF that predicts water retention that is too low. Then the flow model will predict the much faster movement of an infiltration pulse through the soil profile than should happen in reality. In such a case, a water loss will be predicted at some depths for some period of redistribution of the

infiltration pulse while in actuality water will still be accumulating in these horizons. In such a case, the model with this PTF will indeed predict behavior opposite to that observed, and the values of weights from regression-based multimodeling may be negative for this model. Note that the SVD method, which allows negative weights, performed the best.

The list of methods to build the multimodel ensemble in this work is far from exhaustive. For example, principal component analysis (Young, 2002) can be used to replace correlated predictors F_i in Eq. [1] with a smaller number of uncorrelated predictors. The number of independent variables in Eq. [1] can be decreased by using only a subset of models (Kharin and Zwiers, 2002). Various statistics have been proposed to select the "best" subset of input variables (Neter and Wasserman, 1974; Kharin and Zwiers, 2002; Regonda et al., 2006). Young (2002) and Regonda et al. (2006) listed several other methods to combine models. Recent developments include using patterns found in experimental data to adjust the weights of individual forecasts in multimodel predictions (Zheng et al., 2004). There is no reason why a nonlinear combination of individual simulations could not be used in multimodels. Using data mining techniques, such as artificial neural networks, may be beneficial toward that end.

Both the accuracy and, to a lesser extent, the reliability of the SVD multimodel had temporal variations (Fig. 6). The accuracy but not the reliability of the SVD multimodel was worse when the training occurred between Days 135 and 185. Juxtaposing data in Fig. 1 and Fig. 6 shows that the accuracy decreased for training periods that included a period of intensive drying followed by extensive wetting. This was not correctly reflected by simulations based on the Richards flow model. All individual models performed somewhat poorly in a qualitative sense during these time periods, and training was less successful than when the Richards equation was a satisfactory model. One possible remedy would be using different weighting coefficients by introducing an indicator variable for prevailing antecedent soil conditions or boundary conditions. The viability of such an approach should depend on the magnitude of the available data sets. This may present an interesting avenue for future research.

The relatively poor performance of the superensemble method was somewhat expected based on previous experience with climate and hydrologic models (i.e., Regonda et al., 2006). Using the best model was the second best method in this work. The superiority of a multimodel prediction to that of a single model has been observed consistently in meteorological forecasts. This is attributed to the fact that it is difficult or impossible to define the single best model (Hagedorn et al., 2005) if various training and testing periods are considered. Figure 8 shows that there was no generally superior model in our study, and that the model considered the best depends on when the training period started. Hagedorn et al. (2005) noted that a single model may perform better in some situations but that the multimodel approach in the long term will give more reliable simulations.

The premise of superiority of multiple-source simulation systems is based on the hypothesis that "two or more inaccurate but independent predictions of the same future events may be combined in a very specific way to yield predictions that are on average more accurate than either of any of them taken individually" (Thompson, 1977). This hypothesis is supported by some artificial neural networks studies, specifically the group method

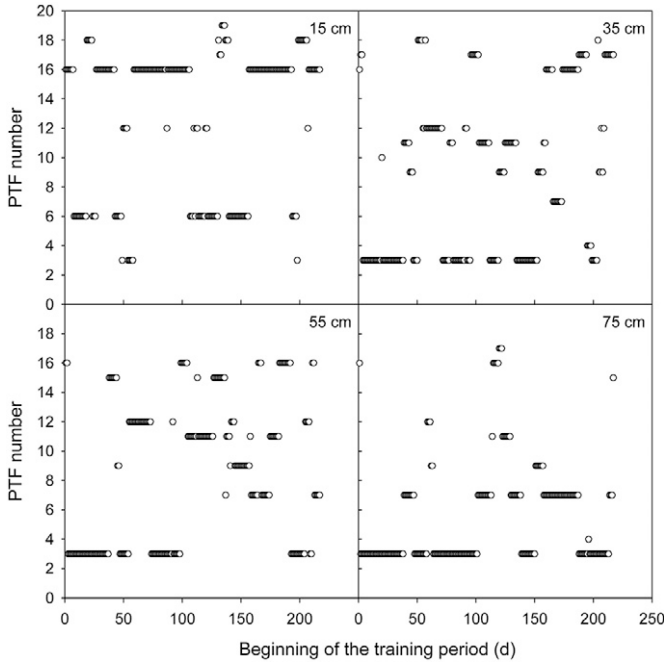

FIG. 8. Pedotransfer functions (PTFs) that provided the most accurate simulations at four depths using a 30-d training period for different days of the beginning of training.

of data handling (e.g., Pachepsky and Rawls, 1999), which uses combinations of individually poor predictions to create predictors with excellent performance. The results of this study seem to further corroborate this hypothesis.

The weights of modeling results with individual PTFs depended on the training period (Fig. 7). This is an expected outcome from the regression analysis when different regression coefficients are obtained if different data subsets are used to develop the regression. Similarly, the training and testing accuracies were also dependent on the training period (Fig. 4). In spite of differences in weight values, however, the overall accuracy of the multimodel depended much more on the method used to combine predictions than on the testing data set (Fig. 4). The selection of the training data set affected the participation of individual models in the multimodeling results, but the overall accuracy and reliability of the multimodel did not vary much. In particular, the accuracy and reliability of the SVD multimodel remained relatively high for all training data sets (Fig. 4).

We include here a comment about terminology. The term *ensemble prediction* is sometimes also used to refer to the integration of the results of many simulation runs but for cases where the initial conditions and parameters or models are randomly perturbed. The term *multimodel ensemble prediction* is used when both different models and their input perturbations are used in individual simulation runs.

In this study, we deliberately made no attempt to discriminate between PTFs or to single out the best PTF or PTFs. First, the multimodel concept does not presume any such analysis since it is agreed from the outset that all models will contribute. Second, based on the data in Fig. 8, the result would depend on the training period. We note that several attempts have been made to relate the performance of PTFs to similarity in geographic regions, to the size of the database used to develop a PTF, to the

homogeneity of this database in terms of measurement methods, or other factors (Pachepsky and Rawls, 2004). All of these attempts appear to have been inconclusive. Our results by no means make a general statement about the quality or reliability of an individual PTF because the results depend on the chosen data set and training period.

The similarity in accuracy between the calibrated Richards model for the layered soil and the multimodel predictions with PTFs suggest the interesting possibility of using soil moisture monitoring data to obtain a more accurate predictive soil water flow model. Monitoring soil moisture and finding weights for the multimodel prediction may well be a very viable approach to simulating field water flow in the vadose zone.

Acknowledgments

The work has been partially supported through the Interagency Agreement RES-02-008, "Model Abstraction Techniques for Soil Water Flow and Transport."

References

Armstrong, J.S. 2001. Combining forecasts. p. 417–439. *In* J.S. Armstrong (ed.) Principles of forecasting: A handbook for researchers and practitioners. Kluwer Acad. Publ., Dordrecht, the Netherlands.

Barnston, A.G., S. Mason, L. Goddard, D.G. Dewitt, and S.E. Zebiak. 2003. Multimodel ensembling in seasonal climate forecasting at IRI. Bull. Am. Meteorol. Soc. 84:1783–1796.

Bates, J.M., and C.W.J. Granger. 1969. The combination of forecasts. Oper. Res. Q. 20:451–468.

Baumer, O.M. 1992. Predicting unsaturated hydraulic parameters. p. 341–354. *In* M.Th. van Genuchten et al. (ed.) Proc. Int. Worksh. on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. U.S. Salinity Lab., Riverside, CA.

Brooks, R.H., and A.J. Corey. 1964. Hydraulic properties of porous media. Hydrol. Pap. 3. Colorado State Univ., Fort Collins.

Burnham, K.P., and D.R. Anderson. 2002. Model selection and inference: A practical information-theoretic approach. Springer-Verlag, New York.

Campbell, G.S., and S. Shiozawa. 1992. Prediction of hydraulic properties of soils using particle size distribution and bulk density data. p. 317–328. *In* M.Th. van Genuchten et al. (ed.) Proc. Int. Worksh. on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. U.S. Salinity Lab., Riverside, CA.

Canarache, A. 1993. Physical-technological maps: A possible product of soil survey for direct use in agriculture. Soil Technol. 6:3–16.

Clemen, R.T. 1989. Combining forecasts: A review and annotated bibliography. Int. J. Forecast. 5:559–583.

Doblas-Reyes, F.J., M. Déqué, and J.-Ph. Piedelievre. 2000. Model and multimodel spread and probabilistic seasonal forecasts in PROVOST. Q. J. R. Meteorol. Soc. 126:2069–2088.

Doblas-Reyes, F.J., R. Hagedorn, and T.N. Palmer. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting. II: Calibration and combination. Tellus Ser. A 57:234–252.

FAO. 1975. Soil map of the world at 1:5,000,000. Vol. 1. Europe. FAO, Rome.

Graham, R.J., A.D.L. Evans, K.R. Mylne, M.S.J. Harrison, and K.B. Robertson. 2000. An assessment of seasonal predictability using atmospheric general circulation models. Q. J. R. Meteorol. Soc. 126:2211–2240.

Guber, A.K., Ya.A. Pachepsky, M.Th. van Genuchten, W.J. Rawls, D. Jacques, J. Simunek, R.E. Cady, and T.J. Nicholson. 2006. Field-scale water flow simulations using ensembles of pedotransfer functions for soil water retention. Vadose Zone J. 5:234–247.

Gupta, S.C., and W.E. Larson. 1979. Estimating soil water retention characteristics from particle-size distribution, organic matter percent, and bulk density. Water Resour. Res. 15:1633–1635.

Hagedorn, R., F.J. Doblas-Reyes, and T.N. Palmer. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting: I. Basic concept. Tellus Ser. A 57:219–233.

Hall, D.G.M., M.J. Reeve, A.I. Thomasson, and V.F. Wright. 1977. Water retention, porosity and density of field soils. Soil Surv. Tech. Monogr. 9. Rothamsted Exp. Stn., Harpenden, UK.

Heimovaara, T.J., and W. Bouten. 1990. A computer controlled 36-channel time domain reflectometry system for monitoring soil water contents. Water Resour. Res. 26:2311–2316.

Hill, C.M. 1998. Methods and guidelines for effective model calibration. Water Resour. Invest. Rep. 98–4005. USGS, Denver, CO.

Jacques, D. 2000. Analysis of water flow and solute transport at the field scale. Ph.D. diss. 454. Faculteit Landbouwkundige en Toegepaste Wetenschappen, Katholieke Univ. Leuven, Leuven, Belgium.

Jacques, D., J. Simunek, A. Timmerman, and J. Feyen. 2001. Calibration of the Richards' and convection–dispersion equations to field-scale water flow and solute transport under rainfall conditions. J. Hydrol. 259:15–31.

Jolliffe, I.N., and D.B. Stephenson. 2003. Forecast verification: A practitioner's guide in atmospheric science. John Wiley & Sons, New York.

Kharin, V., and F.W. Zwiers. 2002. Climate predictions with multimodel ensembles. J. Clim. 15:793–799.

Krishnamurti, T.N., C.M. Kishitawal, Z. Zhang, T. Larow, D. Bachiochi, and E. Williford. 2000. Multimodel ensemble forecasts for weather and seasonal climate. J. Clim. 13:4196–4216.

Mayr, T., and N.J. Jarvis. 1999. Pedotransfer functions to estimate soil water retention parameters for a modified Brooks–Corey type model. Geoderma 91:1–9.

McLachlan, G.J., and T. Krishnan. 1997. The EM algorithm and extensions. John Wiley & Sons, New York.

Neter, J., and W. Wasserman. 1974. Applied linear statistical models: Regression, analysis of variance, and experimental designs. Richard D. Irwin, Homewood, IL.

Pachepsky, Ya.A., A.K. Guber, M.Th. van Genuchten, T.J. Nicholson, R.E. Cady, J. Simunek, and M.G. Schaap. 2006. Model abstraction techniques for soil water flow and transport. NUREG/CR-6884. U.S. Nucl. Regulatory Commiss., Washington, DC.

Pachepsky, Ya.A., and W.J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. Soc. Am. J. 63:1748–1757.

Pachepsky, Ya.A., and W.J. Rawls. 2004. Status of pedotransfer functions. p. vii–xviii. *In* Ya.A. Pachepsky and W. J. Rawls (ed.) Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam.

Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, et al. 2004. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull Am. Meteorol. Soc 85:853–872.

Palmer, T.N., C. Brankovic, and D.S. Richardson. 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. Q. J. R. Meteorol. Soc. 126:2013–2035.

Pavan, V., and F.J. Doblas-Reyes. 2000. Multimodel seasonal hindcasts over the Euro-Atlantic: Skill scores and dynamic features. Clim. Dyn. 16:611–625.

Peng, P., A. Kumar, H. van den Dool, and A.G. Barnston. 2002. An analysis of multimodel ensemble predictions for seasonal climate anomalies. J. Geophys. Res. 107(D23):4710, doi:10.1029/2002JD002712.

Petersen, G.W., R.L. Cunningham, and R.P. Matelski. 1968. Moisture characteristics of Pennsylvania soils: I. Moisture retention as related to texture. Soil Sci. Soc. Am. Proc. 32:271–275.

Poeter, E., and D. Anderson. 2005. Multimodel ranking and inference in ground water modeling. Ground Water 43:597–605.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1992. Numerical recipes in FORTRAN 77: The art of scientific computing. Cambridge Univ. Press, Cambridge, UK.

Raftery, A.E., R. Balabdaoui, T. Gneiting, and M. Polakowski. 2003. Using Bayesian model averaging to calibrate forecast ensembles. Tech. Rep. 440. Dep. of Statistics, Univ. of Washington, Seattle.

Rajagopalan, B., U. Lall, and S.E. Zebiak. 2002. Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. Mon. Weather Rev. 130:1792–1811.

Rajkai, K., and Gy. Várallyay. 1992. Estimating soil water retention from simpler properties by regression techniques. p. 417–426. *In* M.Th. van Genuchten et al. (ed.) Proc. Int. Worksh. on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. U.S. Salinity Lab., Riverside, CA.

Rawls, W.J., and D.L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling. p. 293–299. *In* E.B. Jones and T.J. Ward (ed.) Proc. Symp. Watershed Management in the Eighties, Denver, CO. 30 Apr.–1 May 1985. Am. Soc. Civil Eng., New York.

Rawls, W.J., D.L. Brakensiek, and B. Soni. 1983. Agricultural management ef-

fects on soil water processes. I: Soil water retention and Green–Ampt parameters. Trans. ASAE 26:1747–1752.
Rawls, W.J., D.L. Brakensiek, and K.E. Saxton. 1982. Estimation of soil water properties. Trans. ASAE 25:1316–1320.
Rawls, W.J., D. Giménez, and R. Grossman. 1998. Use of soil texture, bulk density and slope of the water retention curve to predict saturated hydraulic conductivity. Trans. ASAE 41:983–988.
Regonda, S.K., B. Rajagopalan, M. Clark, and E. Zagona. 2006. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resour. Res. 42:W09404, doi:10.1029/2005WR004653.
Saxton, K.E., W.J. Rawls, J.S. Romberger, and R.I. Papendick. 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50:1031–1036.
Schaap, M.G. 2004. Graphic user interfaces for pedotransfer functions. p. 349–356. *In* Ya.A. Pachepsky and W.J. Rawls (ed.) Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam.
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.
Shukla, J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, et al. 2000. Dynamical seasonal prediction. Bull. Am. Meteorol. Soc. 81:2593–2606.
Simunek, J., M. Sejna, and M.Th. van Genuchten. 1998. The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 2.0. IGWMC- TPS- 70. Int. Ground Water Modeling Ctr., Colorado School of Mines, Golden.
Soil Survey Staff. 1999. Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd ed. U.S. Gov. Print. Office, Washington, DC.
Thompson, P.D. 1977. How to improve accuracy by combining independent forecasts. Mon. Weather Rev. 105:228–239.
Tomasella, J., and M.G. Hodnett. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci. 163:190–202.
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.
Varallyay G., K. Rajkai, Ya.A. Pachepsky, and R.A. Shcherbakov. 1982. Mathematical description of soil water retention curve. (In Russian.) Pochvovedenie 4:77–89.
Vereecken, H., J. Maes, J. Feyen, and P. Darius. 1989. Estimating the soil moisture retention characteristics from texture, bulk density and carbon content. Soil Sci. 148:389–403.
Williams, J., P. Ross, and K. Bristow. 1992. Prediction of the Campbell water retention function from texture, structure, and organic matter. p. 427–442. *In* M.Th. van Genuchten et al. (ed.) Proc. Int. Worksh. on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. U.S. Salinity Lab., Riverside, CA.
Wösten, J.H.M., A. Lilly, A. Nemes, and C. Le Bas. 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90:169–185.
Ye, M., S.P. Neuman, and P.D. Meyer. 2004. Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. Water Resour. Res. 40:W05113, doi:10.1029/2003WR002557.
Young, G. 2002. Combining forecasts for superior prediction. p. 107–111. *In* Conf. on Probability and Statistics in the Atmos. Sci., 16th, Orlando, FL. 12–17 Jan. 2002. Am. Meteorol. Soc., Boston, MA.
Yun, W.T., L. Stefanova, and T.N. Krishnamurti. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. J. Clim. 16:3834–3840.
Zheng, X., M. Sugi, and C.S. Frederiksen. 2004. Interannual variability and predictability in an ensemble of climate simulations with the MRI-JMA AGCM. J. Meteorol. Soc. Jpn. 82:1–18.