Data Assimilation with Soil Water Content Sensors and Pedotransfer Functions in Soil Water Flow Modeling

Feng Pan
Dep. of Environmental Science and Policy
Univ. of Maryland
College Park, MD 20742

Current address:
Dep. of Civil & Environmental Engineering
Energy & Geoscience Institute
The Univ. of Utah
Salt Lake City, UT 84112

Yakov Pachepsky*
USDA-ARS
Environmental Microbial and Food Safety Lab.
Beltsville, MD 20705

Diederik Jacques
Institute for Environment, Health, and Safety
Belgian Nuclear Research Centre (SCK•CEN)
BE-2400 Mol Belgium

Andrey Guber
USDA-ARS Environmental Microbial and Food Safety Lab.
Beltsville, MD 20705

Current address:
Dep. of Crop and Soil Sciences
Michigan State Univ.
East Lansing, MI 48824

Robert L. Hill
Dep. of Environmental Science and Policy
Univ. of Maryland
College Park, MD 20742

Soil water flow models are based on simplified assumptions about the mechanisms, processes, and parameters of water retention and flow. That causes errors in soil water flow model predictions. Data assimilation (DA) with the ensemble Kalman filter (EnKF) corrects modeling results based on measured state variables, information on uncertainty in measurement results and uncertainty in modeling results. The objectives of this work were (i) to evaluate pedotransfer functions (PTFs) as a source of data to generate an ensemble of Richards' equation-based models for the EnKF application to the assimilation of soil water content data and (ii) to research how effective assimilation of soil moisture sensor data can be in correcting simulated soil water content profiles in field soil. Data from a field experiment were used in which 60 two-rod time domain reflectometry (TDR) probes were installed in a loamy soil at five depths to monitor the soil water content. The ensemble of models was developed with six PTFs for water retention and four PTFs for the saturated hydraulic conductivity (K_{sat}). Measurements at all five depths and at one or two depths were assimilated. Accounting for the temporal stability of water contents substantially decreased the estimated noise in data. Applicability of the Richards' equation was confirmed by the satisfactory calibration results. In absence of calibration and data assimilation, simulations developed a strong bias caused by the overestimation of K_{sat} from PTFs. Assimilating measurements from a single depth of 15 cm or of 35 cm provided substantial improvements at all other observation depths. An increase in data assimilation frequency improved model performance between the assimilation times. Overall, bringing together developments in pedotransfer functions, temporal stability of soil water patterns, and soil water content sensors can create a new source of data to improve modeling results in soil hydrology and related fields.

Abbreviations: DA, data assimilation; EnKF, ensemble Kalman filter; K_{sat}, saturated hydraulic conductivity; PDF, probability density function; PTF, pedotransfer function; TDR, time domain reflectometry.

A large number of soil water flow and storage models have been developed for applications in hydrology, meteorology, agronomy, contaminant hydrology, and other fields. Each of these models is based on a set of simplified assumptions about the mechanisms, processes, and parameters of water retention and flow, and it is often not possible to predict whether a particular set of assumptions will be applicable for a specific site. Therefore, errors in soil water modeling predictions arise that result from both conceptual uncertainty and the lack of detailed knowledge about model parameters.

Soil water content monitoring data can be used to decrease errors in models. One way to do that is to monitor soil water content for a long period of time and

Soil Sci. Soc. Am. J. 76:829–844
doi:10.2136/sssaj2011.0090
Received 7 Mar. 2011.
*Corresponding author (Yakov.Pachepsky@ars.usda.gov).

to calibrate the model. This is usually a nontrivial task given the high nonlinearity of realistic soil water flow models, layering that may require separate parameter sets and observations within each layer, and the need to accumulate observations of a substantial number of both flood and dry spell events.

Using monitoring data to periodically correct modeling results is a different way to reduce modeling errors. The correction consists in updating simulated values, that is, replacing simulated values of environmental variables with values that are closer to the measured ones. This operation is called DA. It has become a common approach in modeling atmospheric and oceanic systems (Lahoz et al., 2010).

Data assimilation in soil water flow and storage modeling has a substantial history. First applications were focused on modeling water storage in irrigated soils with soil water balance computed for the whole soil profile (Aboitiz et al., 1986; Or and Hanks, 1992). Neutron probe measurements were used to correct the simulated total soil water storage in soil profiles (1.5-m deep in the work of Or and Hanks, 1992). Wendroth et al. (1999) showed that DA in soil water modeling could be efficient if the soil water model includes three layers. A large volume of research was devoted to assimilating remote sensing data on surface soil moisture to infer the profile distribution of soil water contents. Originally, coupled heat transport and water equations were used as the model needing corrections by DA (Entekhabi et al., 1994; Walker et al., 2001a). Later, the soil water flow model given by Richards' equation was used in remote sensing DA (Heathman et al., 2003; Das and Mohanty, 2006). Semiempirical soil water flow and storage models with a small number of vertical compartments were used for coarser spatial scales (Crow and Van Loon, 2006; Huang et al., 2008).

The simplest way of DA is the direct insertion of the measured values of state variables in place of simulated ones. Although this DA method has been applied from time to time (Houser et al., 1998; Walker et al., 2001b; Heathman et al., 2003), it has been recognized that DA-based correction of modeling results should use information on uncertainty in data and uncertainty in modeling results. Simulated values should be changed to the values very close to measured ones if the uncertainty in data is much less than the uncertainty in modeling results. On the other hand, there is no reason to substantially change simulated values if the uncertainty in modeling results is much less than the uncertainty in data. This concept has been formalized by applying the statistical technique called Kalman filter which is a proven data assimilation method for linear dynamics and measurement processes with Gaussian error statistics (Kalman, 1960). This technique has been applied from the very beginning of data assimilation in soil moisture modeling (Aboitiz et al., 1986; Or and Hanks, 1992). As the DA for nonlinear models became of interest, the EnKF was proposed by Evensen (1994) to overcome limitations of Kalman filter. The EnKF is a sequential DA method, which uses an ensemble of model states to represent the error statistics of the model estimation. The idea is to start an ensemble of (many) simulations by varying model parameters, initial state variables, and forcing within feasible ranges. The variation in modeling results within the ensemble at the time of state variable update is used to define the uncertainty in modeling results. Vereecken et al. (2008) noted that the conceptual simplicity, relative ease of implementation, and computational efficiency of the EnKF make the method an attractive option for DA in vadose zone hydrology. The EnKF has been proven to be an efficient approach to correct Richards' equation-based soil flow modeling results of soil water contents by assimilating surface soil moisture (Das and Mohanty, 2006).

Soil moisture DA from sources other than remote sensing of surface soil moisture received little attention so far. At the same time, soil water content or soil matric potential sensors have become the wide-spread source of data on water contents in deep soil layers (Vereecken et al., 2008). Capacitance sensors, for example, have been used in irrigation scheduling (Fares et al., 2006), estimating soil hydraulic properties (Kelleners et al., 2005), evaluating tree water uptake (Schaffer, 1998), upscaling soil water contents (Guber et al., 2009) and many other applications. Examples of soil moisture sensors data assimilation are not numerous, and include the pioneer work of Wendroth et al. (1999) on assimilation of tensiometer data, and assimilation of vadose zone recharge data (Ng et al., 2009).

Selection of the ensemble of models can strongly affect the efficiency of data assimilation with EnKF. Crow and Van Loon (2006) noted that in land data assimilation, relatively little guidance exists concerning strategies for selecting the appropriate magnitude and/or type of introduced model noise. They used the example of coarse-scale soil water model to demonstrate that inappropriate model error assumptions can worsen the performance of a model. In case of Richards' equation-based soil water modeling, feasible ranges of initial soil water contents can be established for a specific case, but establishing a feasible ensemble of soil water flow parameter sets is far from trivial. It was recently proposed to build an ensemble of soil water flow simulations using an ensemble of PTFs (Guber et al., 2006, 2008). The argument went that the accuracy of PTF outside the data collection region is essentially unknown, and the ensemble forecasts offer a way of filtering the predictable from the unpredictable through averaging–the features that are consistent among ensemble members are preserved, while those that are inconsistent are reduced in amplitude. Perhaps more important, the ensemble itself, as a sample from possible forecast outcomes, can be used to estimate the forecast uncertainty and the likely structure of forecast errors (Hamill et al., 2004). Pedotransfer functions were used to adjust the spatial distribution of soil texture and hydraulic properties to match simulated and measured soil moisture when the direct insertion of remotely sensed surface soil water content was used as the DA method (Santanello et al., 2007). However, pedotransfer functions have not been so far used in soil water sensor data assimilation.

The objectives of this work were (a) to evaluate PTFs as a source of data to generate an ensemble of models for the EnKF application to the assimilation of soil water content sensor data

and (b) to research how effective assimilation of soil moisture sensor data can be in correcting simulated soil water content profiles in field soil. A field experimental dataset was used in which the temporal stability of soil water content patterns was observed and used in data assimilation procedure.

MATERIALS AND METHODS

Soil Water Content Monitoring Data

The experimental setup and soil water content data have been previously described by Jacques (2000) and Pachepsky et al. (2005). In brief, the experimental field was located at Bekkevoort, Belgium. It was situated at the bottom of a gentle slope and was covered with a meadow. The soil was classified as Eutric Regosol (FAO, 1975). A trench, 1.2-m deep and 8-m long, was dug at the field site. The grass cover was removed from the experimental area. A plastic sheet to isolate the disturbed trench zone covered one side of the trench. Volumetric water content was measured with TDR. Sixty two-rod TDR probes (25-cm long, 0.5-cm rod diam., 2.5-cm rod spacing) were installed along the 5.5 m of the trench at 12 locations each 50 cm at five depths of 15, 35, 55, 75, and 95 cm (Fig. 1). Soil texture and organic matter content were measured in samples taken where the probes were installed. Grain-size analyses of the sand samples have been performed according to the European standard EN 933-1. Samples were prepared by eliminating carbonates and organic matter. For the particles larger than 50 μm, a standard sieving was used with mesh sizes of 100, 250, 500, and 1000 μm. For the particles smaller than 50 μm, a dispersing agent was added first after which the solid/water mix was put in a suspension cylinder for determination of the fine particles with class boundaries of 2, 11, and 22 μm. The organic matter content was determined using the rapid dichromate oxidation method adapted from the Walkley–Black procedure. Soil texture was sandy loam at depths of 15, 35, and 55 cm, and loam at depths of 75 and 95 cm. One measurement cycle for all TDR-probes took approximately 35 min, and the time difference between two measurements for the same probe was 2 h. After all devices were installed, the trench was filled. Rainfall was continuously measured at the site with a rainfall recorder (200 cm^2) with a floated pen system on a paper (0.1 mm interval, rotation speed 1 cm h^{-1}). Other meteorological parameters were obtained from the station 3 km from the site. A thin layer of gravel (1–2 cm) was evenly distributed on the study area: (i) to decrease the erosive effect of the rain impact on the bare soil surface, (ii) to minimize the evaporation from the soil surface, and (iii) to decrease the growth of weed on the experimental plot. Weeds were regularly removed from the site during the summer. Field measurements started on 11 Mar. 1998 (Day 0) and finished on 31 Mar. 1999 (Day 384). A site-specific TDR calibration (Jacques, 2000) was used.

Ensemble Kalman Filter: Theory and Application

The Kalman filter is an implementation of the Bayesian update method. Given a probability density function (PDF) of the state of the modeled system (the prior) and the probability distribution function of data, the Bayes theorem is used to obtain the PDF after the data has been taken into account (the posterior). The Bayesian update incorporates new data when they become available, and model advances in time from one update to another. The following description is based on the work of Mandell (2007).

The Kalman filter relies on normal distributions of data and modeling results. Let the model for any simulated time generate N state variables $x_1, x_2, \ldots x_{n..}$ The PDF $p(x)$ of the vector of simulation results $\mathbf{x} = \{x_1, x_2, \ldots x_{n.}\}$ is

$$p(\mathbf{x}) = A_1 \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{Q}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \qquad [1]$$

where $\boldsymbol{\mu}$ is the vector of mean values of variables x_1, x_2,, x_n, $\mathbf{Q}$ is the covariance matrix, A_1 as well as A_2, A_3, and A_4 in equations below, are scaling multipliers to have the integral of probability distribution function equal to one. The function $p(\mathbf{x})$ is the prior probability distribution on the moment the state has

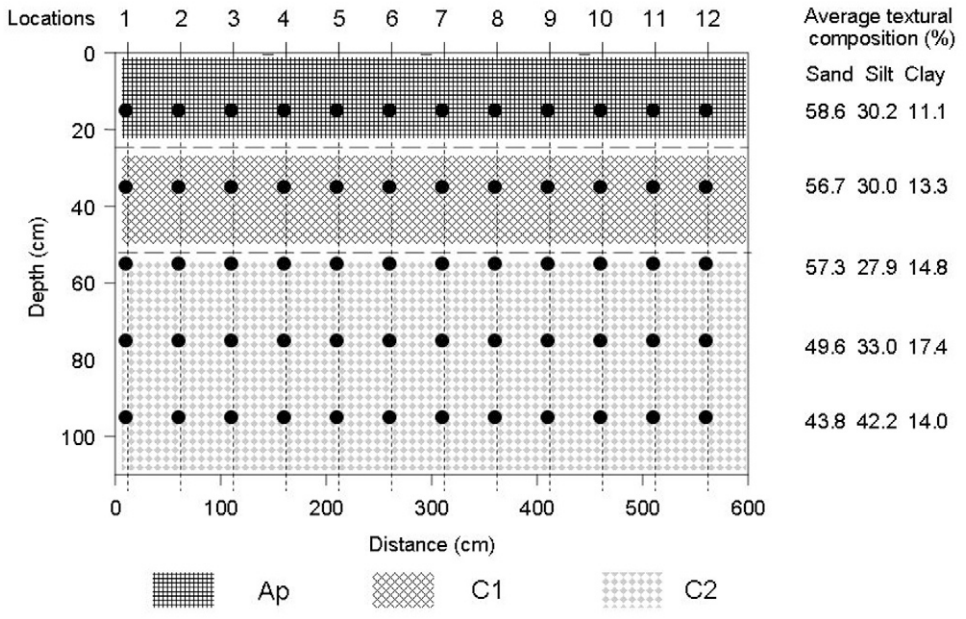

Fig. 1. Time domain reflectometry probe placement (dot) at the trench wall. Locations 1 through 12 denote 12 positions along the trench where sets of five sensors–one for each measurement depths were installed. Filled rectangles show Ap, C1, and C2 horizons top to bottom. Dashed lines show the average position of the horizon boundary, and white bands show the observed range of horizon boundary depths. Average values of clay, silt, and sand content are given for the probe installation depths. Adopted from (Pachepsky et al., 2005).

to be updated to account for data. The vector of data values **d** is also assumed to be normally distributed with the mean $\bar{\mathbf{d}}$ and covariance matrix **R.** It is assumed that the mean data vector $\bar{\mathbf{d}}$ is related to state variables **x** via matrix **H** as $\bar{\mathbf{d}} = \mathbf{Hx}$. The value **Hx** is what the value of the data would be for the state **x** in the absence of data errors. Then the probability density $p(\mathbf{d}|\mathbf{x})$ of the data **d** conditional of the system state **x**, is

$$p(\mathbf{d}\,|\,\mathbf{x}) = A_2 \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{Hx})^T\,\mathbf{R}^{-1}(\mathbf{d}-\mathbf{Hx})\right) \quad [2]$$

For the update purposes, one needs the probability density of states conditioned on data $p(\mathbf{x}\,|\,\mathbf{d})$ rather than the probability density of data conditioned on states $p(\mathbf{d}\,|\,\mathbf{x})$. The conversion of $p(\mathbf{d}\,|\,\mathbf{x})$ to $p(\mathbf{x}\,|\,d)$ can be done using the Bayes theorem

$$p(\mathbf{x}\,|\,\mathbf{d}) = A_3 p(\mathbf{d}\,|\,\mathbf{x}) p(\mathbf{x}) \quad [3]$$

States conditioned on data, that is, **x|d,** are posterior states, they are referred below as $\mathbf{x}^p$. When [1] and [2] are used to compute the right-hand side of [3], the expression for $p(\mathbf{x}^p)$ is obtained in the form:

$$p(\mathbf{x}^p) = A_3 \exp\left[-\frac{1}{2}(\mathbf{x}^p-\boldsymbol{\mu}^p)^T\,\mathbf{Q}^{p-1}(\mathbf{x}^p-\boldsymbol{\mu}^p)\right] \quad [4]$$

The posterior mean $\boldsymbol{\mu}^p$ and posterior covariance $\mathbf{Q}^p$ in Eq. [4] are given by the Kalman update formulas:

$$\begin{aligned}\boldsymbol{\mu}^p &= \boldsymbol{\mu} + \mathbf{K}(\mathbf{d} - \mathbf{H}\boldsymbol{\mu}) \\ \mathbf{Q}^p &= (\mathbf{I} - \mathbf{KH})\mathbf{Q}\end{aligned} \quad [5]$$

where

$$\mathbf{K} = \mathbf{QH}^\mathbf{T}(\mathbf{HQH}^\mathbf{T} + \mathbf{R})^{-1} \quad [6]$$

is the Kalman update matrix. The Kalman update changes state variables taking into account (i) data available at the moment when predictions have been obtained, (ii) the accuracy of those data, and (iii) variability of state variables. One important feature of the Kalman filter is that the number of elements (measurements) in the data vector **d** is usually much smaller than the number of state variables– elements of the vector **x**.

The EnKF has been developed to overcome the difficulty of using the original Kalman filter in cases when the dependence of the covariance matrix **Q** on time is difficult to find. The EnKF estimates the covariance matrix as the sample covariance computed from the ensemble simulation results. The ensemble is composed from randomly generated equiprobable realizations of the studied model. The randomness may apply to initial conditions, model parameters, and boundary conditions or forcing.

Let the ensemble consists of N models and each model predicts n state variables. Let the predictions of the ith model form the vector $\boldsymbol{x}_i$ that has n elements $\boldsymbol{x}_{i,}$ $i = 1,2,...n$, which are predicted values of state variables. The $n \times N$ matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N]$ is the prior ensemble. The goal is to correct the predictions at each of preset update times by changing this matrix to the posterior ensemble $\mathbf{X}^p = [\mathbf{x}_1^p, \mathbf{x}_2^p, ...\mathbf{x}_N^p]$. It is assumed that the data form the vector **d** that has m elements. The vector ε is the random error in data characterized by the $m \times m$ error covariance matrix **R.**

The EnKF update consists of four basic steps.

1. Find the $n \times N$ covariance matrix **C** of ensemble predictions $\mathbf{x}_i$

2. Generate representative random data separately for each ensemble member: $\mathbf{d}_1 = \mathbf{d} + \varepsilon_1$, $\mathbf{d}_2 = \mathbf{d} + \varepsilon_2$, ..., $\mathbf{d}_N = \mathbf{d} + \varepsilon_N$, where the random vector ε belongs to the n-dimensional normal distribution $N(0,R)$.

3. Collect the random data in the $m \times N$ matrix $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ...\mathbf{d}_N]$

4. Find the corrected predictions as:

$$\mathbf{X}^p = \mathbf{X} + \mathbf{K}(\mathbf{D} - \mathbf{Hx}) \quad [7]$$

where the Kalman gain matrix K relates the variability in predictions and the data accuracy and is estimated as

$$\mathbf{K} = \mathbf{CH}^\mathbf{T}(\mathbf{HCH}^\mathbf{T} + \mathbf{R})^{-1} \quad [8]$$

The one-dimensional case gives a general feel of how the ensemble Kalman filter works. Consider the case $n = 1$ and $m = 1$ when there is only one model-predicted state variable and its value is measured. All matrices then will become scalars, and **H** will be equal to 1. Let $x_i = \boldsymbol{\mu} + \xi_i$, ξ_i belongs to $N(0, \sigma_x^2)$, and $d_i = d + \varepsilon_i$, ε_i belongs to $N(0, \sigma_d^2)$. The gain K will be

$$K = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2} \quad [9]$$

and

$$x_i^p = (1-K)(\mu + \xi_i) + K(d + \varepsilon_i) \quad [10]$$

Values of K are between 0 and 1. The value of x_i^p is close to x_i when K is close to zero, that is, $\sigma_d^2 >> \sigma_x^2$ and accuracy in data is much lower compared with the variability in predictions. On the contrary, the value of x_i^p is close to d_i when K is close to one, that is, $\sigma_x^2 >> \sigma_d^2$ and accuracy in data is much higher than the variability in predictions.

In the application of the EnKF in this work, state variables were water contents at five depths, and therefore n was equal to five. The data vector varied in its size from one (assimilation from only one measurement depth, $m = 1$) to five (assimilation of measurements from all depths, $m = 5$). Since both model results and measurements were soil water contents, the matrix **H** had some diagonal elements h_{ii} (1 = 1,2...,5) equal to 1 and all other elements equal to zero. For example, only $h_{22} = 1$ and $h_{33} = 1$ if the measurements from the second and the third depth were used, and only $h_{11} = 1$ and $h_{44} = 1$ if measurements from the first and fourth depth were used. Computation of data errors and model errors is discussed below.

Soil Water Flow Model

The one-dimensional vertical soil water flow at the Bekkevoort experimental site was simulated with the Richards' equation

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z}+1\right)\right] \quad [11]$$

where θ is the soil water content [$L^3\ L^{-3}$]; h is the matric potential [L]; K is the hydraulic conductivity [$L\ T^{-1}$]; z is the vertical axis directed upward [L]; t is the time [T]. Soil water retention was described using the van Genuchten equation (van Genuchten, 1980):

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

where θ_s, θ_r are saturated and residual soil water content [$L^3\ L^{-3}$]; α [L^{-1}], n, m are van Genuchten water retention parameters. The hydraulic conductivity was computed from the van Genuchten–Mualem equation (van Genuchten, 1980):

$$K = K_{\text{sat}}\left(\frac{\theta-\theta_r}{\theta_s-\theta_r}\right)^l\left\{1-\left[1-\left(\frac{\theta-\theta_r}{\theta_s-\theta_r}\right)^{1/m}\right]^m\right\}^2 \quad [13]$$

where K_{sat} is saturated hydraulic conductivity [$L\ T^{-1}$], l is an empirical shape-defining parameter. The value of the parameter m was set to $1-1/n$.

Equation [11] was solved numerically using the HYDRUS 1D software (Šimůnek et al., 2008). The atmospheric boundary with daily rainfall and evapotranspiration was set as the top boundary condition, and the free drainage boundary condition was set as the bottom boundary condition. The pressure head profile calculated from measured soil water content based on the van Genuchten equation was set as the initial condition. Predicted and updated state variables were water contents at five measurement depths averaged across the 12 observation locations at the beginning of the day of update.

Pedotransfer Functions to Develop the Ensemble of Models

Pedotransfer functions developed from large databases were used to generate parameters in the van Genuchten–Mualem parameterization of soil hydraulic properties in variably saturated soils (Eq. [12] and [13]). Parameters of the water retention function (Eq. [12]) were found from the six pedotransfer functions (Appendix) developed from the European continental database HYPRES (Wösten et al., 1999), subsets of the U.S. nationwide database (Gupta and Larson, 1979; Rawls et al., 1983), the nationwide Brazilian dataset (Tomasella and Hodnett, 1998), and the large national Hungarian database in which sandy loam and loam soils were well represented (Rajkai and Varallyay, 1992). The pedotransfer equations of water retention parameters are described in details in the Appendix.

Four sets of K_{sat} values were used to create ensembles of models in this study (Appendix). The ensemble of 24 models (6 PTFs of water retention × 4 K_{sat} PTFs) was applied in soil moisture data assimilation with EnKF.

Temporal Stability of Water Contents and Data Error Estimates

The random error in data has to be characterized to apply the Kalman update method. In this work, the data are the average values of water contents across the trench at each of five observation depths. The observed time series of soil water contents were previously analyzed in the work of Pachepsky et al. (2005). Substantial temporal stability was found that manifested itself in the similarity of soil water content time series shapes in different locations at the same depth, and shifts of the time series graphs relative to each other along the water content axis (Fig. 2). Because the time series at the same depths were correlated, the "naive" computation of the covariance matrix of data errors D under the assumption of independence of data in different locations at the same depth could result in large inaccuracies (Wigley et al., 1984) since correlated observations result in inflated type 1 errors (Quinn and Keough, 2002). Therefore, the statistical model of the data was assumed in the form (Jacques et al., 2001):

$$\theta_{i,j}(t) = \mu_i(t) + b_{i,j} + \eta_{i,j} \quad [14]$$

where i is the subscript to denote depth, i = 15, 35, 55, 75, 95 cm, j is the subscript to denote location across the trench, j = 1,2,...,12, μ_i is the average water content at the depth "i", $b_{i,j}$ is the bias of the measurement in location j at the depth "i" relative to the average water content at this depth, and η_{ij} is the random component that is used to define the covariance matrix. The bias values were derived by fitting Eq. [14] to the whole observed time series (Table 1). Inspection of the Table 1 shows that the spatial distribution of the bias values is not random; zones of negative and positive bias can be delineated in soil 2D cross-section along the studied transect.

Study Design

Four groups of questions have been addressed in series of computations.

1. Are PTF-based models applicable at the site as is, without any calibration or data assimilation?

2. Can Richards' equation be calibrated to mimic the water flow at the site? Is the Richards' equation applicable?

3. Can DA with measurements at one or two depths result in a satisfactory reproduction of water content time series at other depths? Which depths are more efficient for soil water content DA? How does the DA time interval affect the overall accuracy of simulations?

4. Does DA improve results of simulations with calibrated models?

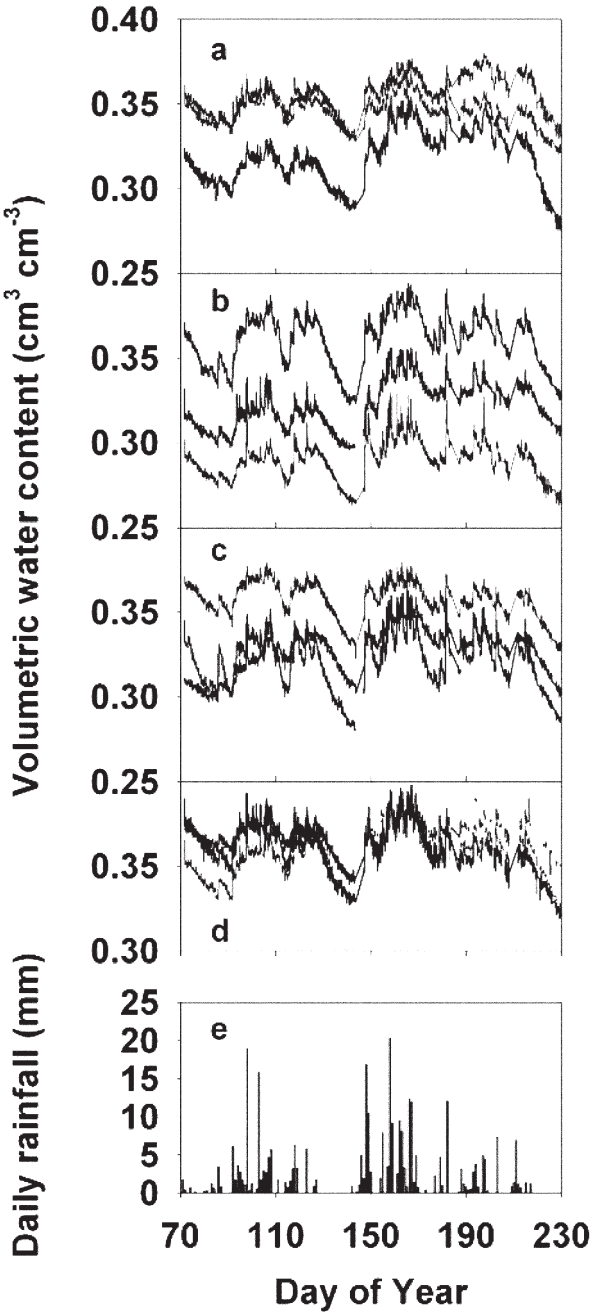

Fig. 2. Time series of time domain reflectometry (TDR)–(a–d) measured water contents at the 15-cm depth and (e) precipitation. Location numbers top to bottom: (a) 2, 3, 1; (b) 5, 6, 4; (c) 7, 8, 9; and (d) 11, 12, 10. Location numbering is shown in Fig. 1. Adopted from (Pachepsky et al., 2005).

To address these questions the ensemble of 24 models was run with soil water retention and saturated hydraulic conductivity parameters estimated with PTFs. Then, each of the water retention PTF was used, and the saturated hydraulic conductivity values for five depths were calibrated with the observations over 30-d observation period from Day 70 to 99. Next, the ensemble of 24 models was run with both soil water retention and saturated hydraulic conductivity parameters estimated with PTFs, and DA was performed with the ensemble Kalman filter algorithm (Eq. [7] and [8]) for daily, weekly, or biweekly updates. And finally, the ensemble of six PTFs with calibrated saturated hydraulic conductivities was run, and the ensemble Kalman filter was applied with weekly updates[1]. The Shapiro–Wilk Test was used to determine if the simulated water contents and data errors were normally distributed at each depth on each assimilation date. The accuracy of simulations was characterized using the root mean squared error (RMSE) values computed as

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N}\left(\theta_i^{(m)} - \theta_i^{(s)}\right)^2 / N} \qquad [15]$$

where N is the number of simulated days, $\theta_i^{(m)}$ and $\theta_i^{(s)}$ are measured and simulated volumetric soil water contents at noon on the day "i".

Calibration of Models in the Ensemble

The Richards' Eq. [11] with the van Genuchten–Mualem hydraulic property models Eq. [12] and [13] was calibrated using the inverse solution option in the HYDRUS 1D software based on the Marquardt–Levenberg algorithm (Šimůnek et al., 2008). The saturated hydraulic conductivity values K_{sat} were subject to calibration separately for each of water retention PTFs; parameters of the water retention equation computed with water retention PTFs were not calibrated. The K_{sat} values were calibrated within depth intervals 0 to 25 cm, 25 to 45 cm, 45 to 65 cm, 65 to 85 cm, and >85 cm.

RESULTS

Applicability of Pedotransfer Function-based Models at the Site without Calibration or Data Assimilation

Ensemble simulations between Day 100 (10 Apr. 1998) and Day 247 (4 Sept. 1998) are summarized in Fig. 3. The PTF-based models appear to be incapable to simulate water flow at the site. Much more water is lost from the soil profile between the rainfall periods in simulations compared to measurements. Guber et al. (2009) observed a similar performance of the ensemble of PTF-based models used without calibration at this site.

Accuracy of Calibrated Models

Calibration of the saturated hydraulic conductivity values led to the successful simulation of water contents at all five depths (Fig. 4). The calibrated Richards' equation was an adequate model to predict soil water flow at the site, at least for precipitation and evaporation encountered during the observation period. Table 2 lists the calibrated K_{sat} values for each of six water retention PTFs. The calibrated K_{sat} varied among the six PTFs, indicating that the choice of the water retention PTFs affects the values of calibrated K_{sat} to some extent (Table 2). Comparison of calibrated K_{sat} with with PTF-based K_{sat} (Appendix) shows that all calibrated values of hydraulic conductivity were substantially smaller than the values predicted with pedotransfer functions.

[1] The FORTRAN code is available on request from the corresponding author.

Data Assimilation with the Noncalibrated Ensemble

Selected data assimilation results are shown in Fig. 5. Data assimilation provided an excellent update of weekly simulation results when the data from all depths were assimilated (Fig. 5a). Inspection of graphs in Fig. 5b and 5c shows that assimilation of measurements from the depth of 15 cm resulted in the same accuracy as assimilation of data from all depths and assimilation of measurements from the depth of 95 cm resulted in relatively large errors in the top of the profile. While the update was satisfactory, the simulations between update times deviated from measurements since parameters of the model were not changed.

The systematic overview of errors in simulations with the ensemble Kalman filter data assimilation is presented in Table 3. The largest simulation errors are found at the depths of 15 and 35 cm, the smallest at the depths of 75 and 95 cm. This happens because the magnitudes and rates of water content changes are much larger at the depth of 15 cm than at the depth of 95 cm. Therefore the deviations of ensemble simulations from measurements during the week between assimilations are much larger in the near-surface soil layers at depths of 15 and 35 cm.

Assimilation of the data from the depth of 35 cm resulted in an RMSE value which was the same or better than the one in the case of assimilation of the data from all depths. With the assimilation of data from only one depth, the accuracy at all depths generally decreased as the assimilation depths increased (Table 3). Assimilation of the data from the depth of 95 cm lead to the worst results in terms of the RMSE. Figure 5c shows that the errors stemmed from discrepancies between updates and measured water contents at shallow depths. The soil water dynamics observed at the depths of 75 and 95 cm did not capture changes occurring at smaller depths.

Interestingly, the simulation accuracy at the depth of 95 cm between the assimilation times was better when the data were assimilated from the depth of 15 cm as compared with assimilation from the depth of 95 cm. This probably happened because the errors caused with assimilation of data from 95 cm translated into substantial errors across the whole profile including the 95-cm depth during the week between updates.

Adding a second assimilation depth generally improved the accuracy of simulations at all depths in most cases, but could decrease the accuracy of simulations if the bottom measurement depths (75 or 95 cm) were added to the top measurement depths (15 or 35 cm) (Table 3). The smallest RMSEs of the water content simulations were obtained after the assimilation of data from (a) the 15-cm depth, (b) the 35-cm depth, (c) from two depths of 15 and 35 cm, and (d) from two depths of 15 and 55 cm (Table 3). However, assimilation of data from other depths was only marginally worse in terms of RMSE values (Table 3).

Biweekly data assimilation has led to the general increase of the simulation RMSE (Table 3). Similarly to the weekly assimilation, smaller RMSE were found when the data from top observation depths of 15 and 35 cm were assimilated. The difference in RMSE between the assimilation of data from the 15 cm depths and the assimilation from other depths was smaller than in case of weekly updates. For example, assimilations of data from 15 cm and from 95 cm lead to the simulation RMSEs of 0.0369 cm^3 cm^{-3} and 0.0401 cm^3 cm^{-3}, respectively, with biweekly updates, and to the simulation RMSEs of 0.0307 cm^3 cm^{-3} and 0.0358 cm^3 cm^{-3}, respectively, with weekly updates.

Table 1. The bias values of the measurements in 12 locations at the five depths with respect to the average water content at each depth.

Time domain reflectometry location	Bias values at the depths of				
	15 cm	35 cm	55 cm	75 cm	95 cm
1	–0.0191	0.0096	–0.0315	–0.0160	0.0152
2	0.0192	0.0103	–0.0039	0.0036	0.0184
3	0.0063	0.0104	–0.0146	–0.0104	0.0058
4	–0.0493	–0.0131	–0.0117	–0.0372	0.0164
5	0.0195	–0.0084	0.0069	0.0160	–0.0054
6	–0.0200	0.0101	0.0003	0.0150	0.0203
7	0.0111	–0.0430	0.0021	–0.0028	–0.0707
8	–0.0167	–0.0214	0.0071	0.0253	0.0083
9	–0.0177	–0.0165	0.0105	0.0009	0.0034
10	0.0149	0.0138	0.0085	0.0074	–0.0010
11	0.0302	0.0257	0.0029	0.0239	0.0041
12	0.0216	0.0223	0.0233	–0.0257	–0.0148

Results of daily data assimilation are shown in Fig. 6. The daily update prevents the development of the simulation bias which has been well pronounced with weekly, and even more so, in biweekly updates (Fig. 5). Using the data from only one depth corrects results throughout the profile in case of daily updates as in a case of less frequent updates. However, using more than one sensor seems to be beneficial, since the use of only one sensor from the 15-cm depth leads to the exaggeration of water content dynamics at larger depths (Fig. 6b), and the use of the sensor from the 90-cm depth does not properly correct the simulated dynamics at 15- and 35-cm depths. The RMSE values for daily assimilation are shown in Table 3. They are substantially up to 12 times less than in the case of weekly simulations. The best overall result has been achieved when all five sensors have been used. The next best overall results have been obtained with pairs of sensors from 15 and 55 cm, and from 35 and 95 cm.

Data Assimilation with the Calibrated Ensemble

Results of data assimilation with calibrated models are summarized in Table 4. They are compared with results without calibration in the same table. Using calibrated models in the ensemble in case of weekly assimilation has resulted in much better overall accuracy as compared with the assimilation with noncalibrated models as the comparison of Tables 3 and 4 shows. The daily assimilation update without calibration, however, resulted in better accuracy than the weekly data assimilation with calibrated models.

Statistical Properties of Ensemble Simulations and Measurements

We note that both simulated water contents and data errors were mostly normally distributed. More than 99% of

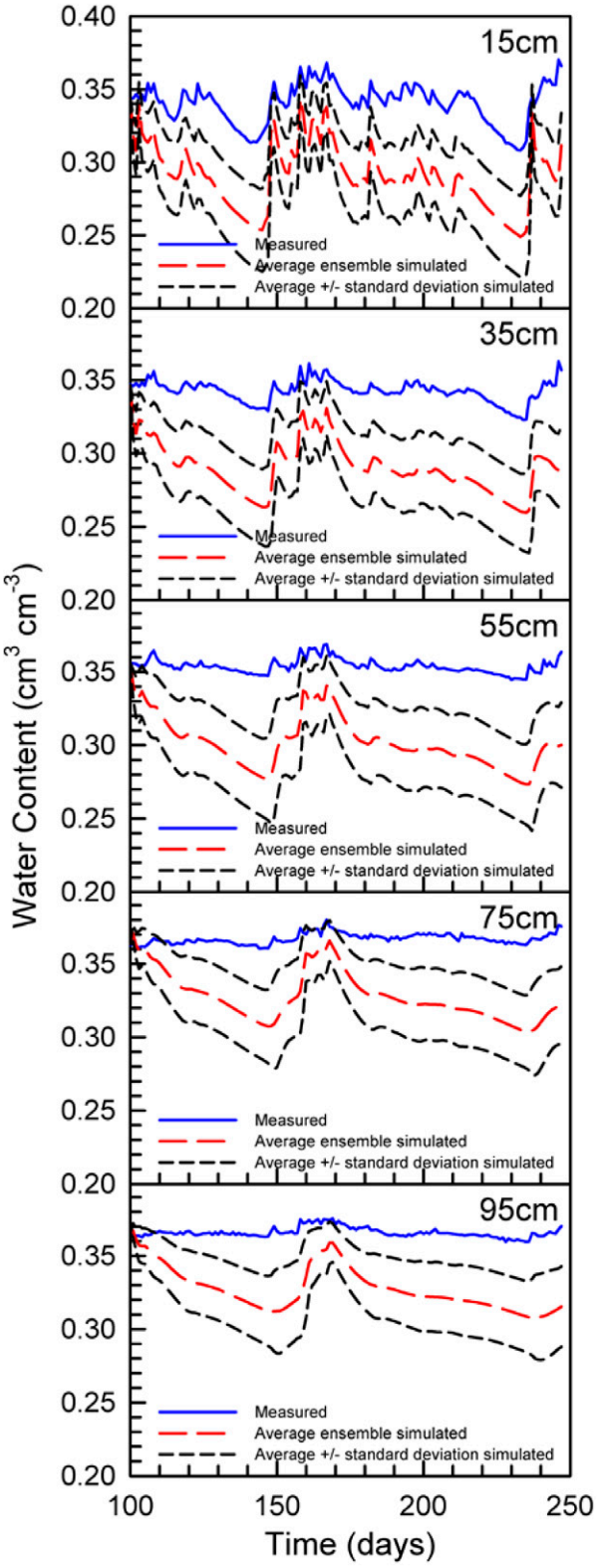

Fig. 3. Comparison of measured and ensemble-simulated soil water contents. Neither calibration nor data assimilation has been applied.

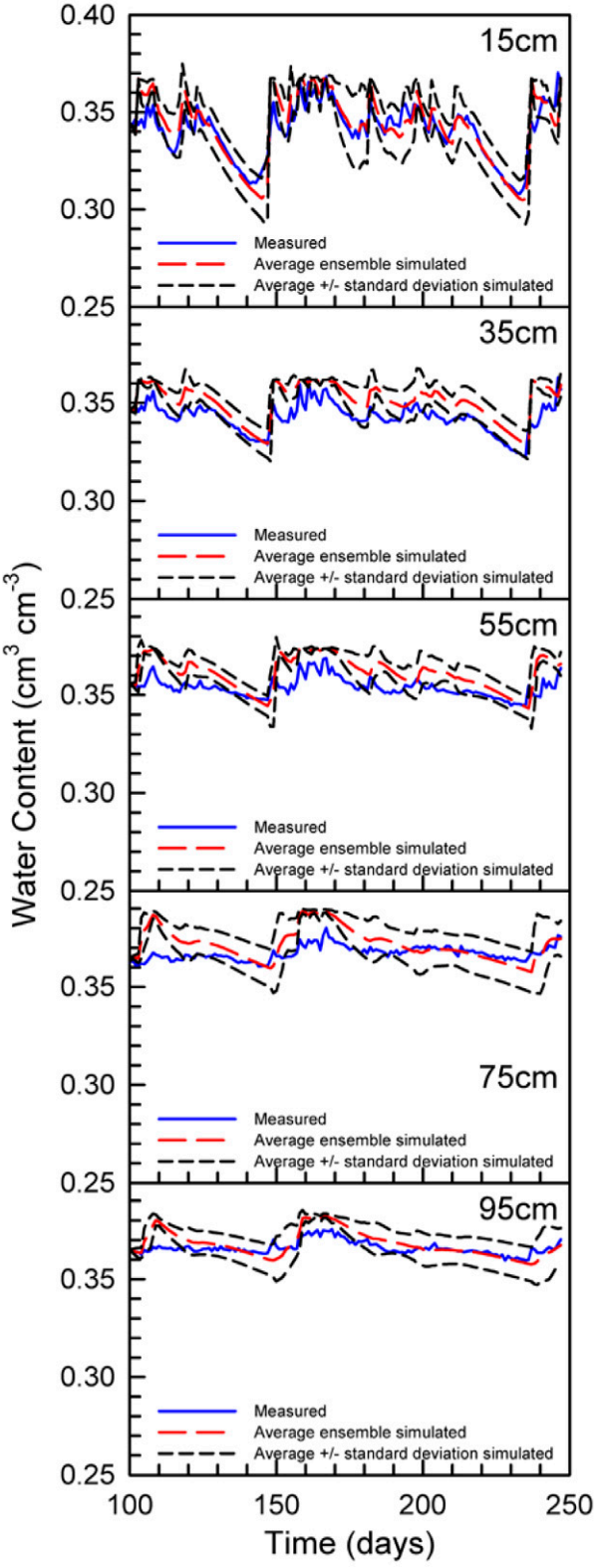

Fig. 4. Comparison of measured and simulated soil water contents. Calibration of saturated hydraulic conductivity has been applied.

data sets passed the normality test at the significance level of 0.01, indicating that the Kalman filter assumption of normal distributions was met for priors and for data. The absence of systematic trends in data errors was also assessed from the inspection of the correlation between the data errors at different depths (Table 5). Correlation coefficients between the data at different depths varied widely among the assimilation dates.

DISCUSSION

Considerable bias has been encountered in the results water flow simulations with the ensemble of pedotransfer functions for

Table 2. Calibrated values of the saturated hydraulic conductivity (K_{sat}) from simulations with six different pedotransfer functions for water retention.

Pedotransfer function no.	Pedotranfer function source	Model	Calibrated K_{sat} (cm d^{-1}) at depths of				
			15 cm	35 cm	55 cm	75 cm	95 cm
1	Wösten et al. (1999)	VG†	4.95	0.93	0.69	2.21	2.56
2	Wösten et al. (1999)	VG†	4.71	0.40	0.68	2.41	1.02
3	Tomasella and Hodnett (1998)	WH→VG‡	3.89	1.55	1.22	1.05	1.35
4	Gupta and Larson (1979)	WH→VG‡	2.44	0.39	0.50	0.95	0.85
5	Rajkai and Varallyay (1992)	WH→VG‡	3.51	1.14	0.14	1.95	2.22
6	Rawls et al. (1983)	WH→VG‡	1.21	0.16	0.47	0.89	0.94

† Parameters of the van Genuchten equation are estimated with the pedotransfer function.

‡ Water contents at specific pressure heads are estimated, and then the van Genuchten equation is fitted to the estimates.

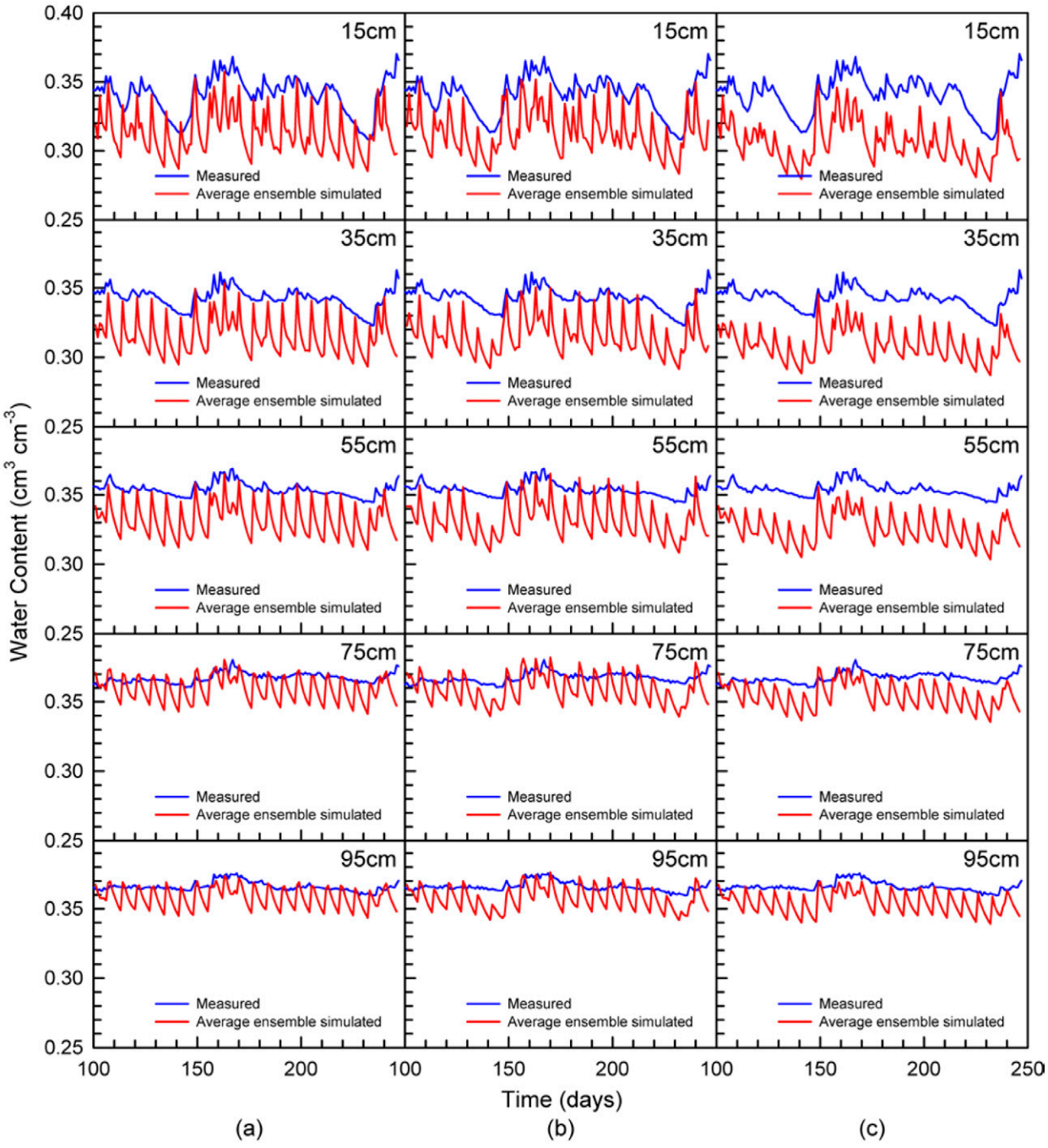

Fig. 5. Selected results of weekly ensemble simulations update; (a) update with assimilation of data from sensors from all five depths, (b) update with assimilation of data from sensors at the 15-cm depth, and (c) update with assimilation of data from the sensors at the 95-cm depth.

Table 3. Root-mean-squared errors of soil water content simulations with weekly, biweekly, and daily data assimilation.

Depth of sensors	Weekly data assimilation					Biweekly data assimilation					Daily data assimilation				
	RMSE (volume %) at the depths of					RMSE (volume %) at the depths of					RMSE (volume %) at the depths of				
	15 cm	35 cm	55 cm	75 cm	95 cm	15 cm	35 cm	55 cm	75 cm	95 cm	15 cm	35 cm	55 cm	75 cm	95 cm
cm															
15,35,55,75,95	3.12	2.98	2.47	1.17	1.06	3.72	3.66	3.27	1.95	1.79	0.52	0.35	0.27	0.45	0.23
15	3.07	3	2.53	1.26	1.17	3.69	3.66	3.3	1.99	1.85	0.41	0.79	0.89	0.9	0.68
35	3.09	2.97	2.45	1.17	1.06	3.7	3.64	3.26	1.95	1.79	0.76	0.45	0.48	0.95	0.69
55	3.18	3.02	2.49	1.21	1.1	3.75	3.68	3.28	1.97	1.8	1.21	0.83	0.31	0.73	0.48
75	3.52	3.39	2.9	1.47	1.34	3.97	3.91	3.54	2.15	1.97	2.43	2.23	1.71	0.31	0.41
95	3.58	3.44	2.97	1.53	1.35	4.01	3.93	3.58	2.19	1.98	2.47	2.22	1.71	0.42	0.29
15, 35	3.06	2.95	2.45	1.2	1.09	3.69	3.65	3.26	1.95	1.8	0.39	0.45	0.6	0.96	0.68
15, 55	3.07	2.96	2.44	1.18	1.07	3.7	3.65	3.26	1.94	1.79	0.56	0.51	0.26	0.74	0.46
15, 75	3.13	3.03	2.54	1.21	1.12	3.73	3.69	3.31	1.98	1.83	0.46	0.82	0.76	0.39	0.3
15, 95	3.11	3.02	2.53	1.22	1.11	3.74	3.69	3.32	1.98	1.82	0.45	0.78	0.7	0.55	0.23
35, 55	3.14	2.97	2.45	1.19	1.07	3.72	3.65	3.26	1.95	1.79	0.9	0.43	0.21	0.81	0.54
35, 75	3.13	2.98	2.47	1.17	1.07	3.73	3.67	3.29	1.96	1.8	0.7	0.46	0.48	0.49	0.3
35, 95	3.14	2.99	2.48	1.19	1.07	3.74	3.68	3.3	1.97	1.8	0.71	0.47	0.44	0.64	0.32
55, 75	3.18	3.02	2.5	1.19	1.09	3.76	3.69	3.29	1.97	1.8	1.18	0.85	0.41	0.41	0.21
55, 95	3.18	3.02	2.5	1.2	1.08	3.76	3.69	3.29	1.97	1.8	1.17	0.85	0.39	0.56	0.23
75, 95	3.53	3.38	2.89	1.46	1.31	3.98	3.91	3.54	2.15	1.96	2.38	2.14	1.63	0.28	0.31

our research site. Simulated water content values declined much faster than measurement (Fig. 1). The DA updates were bringing ensemble simulated water contents closer to measured, but the divergence between simulations and measurements occurred after each update. The reason for the divergence was the large difference between PTF-estimated and actual hydraulic conductivity. The large estimated hydraulic conductivity K_{sat} led to the fast emptying of the soil profile in simulations. The difference between calibrated and PTF predicted values of K_{sat} may be related to the fact that the K_{sat} pedotransfer functions were developed with the data from small soil samples (e.g., Rawls et al., 1998). It has been observed that K_{sat} may decrease with increasing measurement scale (e.g., Mallants et al., 1997). Another reason can be that we adopted the mean values of K_{sat} measured or fitted from a large dataset as the PTF-based K_{sat} values in this study and their standard deviations are large (Schaap and Leij, 1998; Carsel and Parrish, 1988). Yet another possible explanation can be that we observed and simulated mostly unsaturated flow, and in the van Genuchten–Mualem model (Eq. [13]), the effect of K_{sat} on the unsaturated hydraulic conductivity depends on the value of the tortuosity parameter l. We used the generic value of l = 0.5 in all simulations (van Genuchten, 1980), but this value was found to be both positive and negative and to vary in a wide range (Schaap and Leij, 2000). Values of l smaller than 0.5 increase the value of unsaturated hydraulic conductivity, and if the values of l in soil at the site were smaller than values of l in soils in experiments used to derive the K_{sat} PTF, relatively small K_{sat} would be sufficient to fit the unsaturated hydraulic conductivity in the observed ranges of water contents. Also, calibrated K_{sat} should reflect the set of van Genuchten water retention parameters which probably are different of van Genuchten parameters of soils used to develop the K_{sat} PTFs. We note that Jacques et al. (2002) who calibrated both water retention and hydraulic conductivity parameters for the soil of our study have found values of K_{sat} between 1 cm d^{-1} and 4 cm d^{-1} that is close to values we have obtained.

The observed bias in modeling results should affect the DA efficiency, since the data assimilation procedures, including EnKF, are developed assuming random errors both in data and in simulations; the presence of systematic errors in modeling results, however, is a common occurrence that data assimilation encounters (Dee, 2005). Such bias may arise not only from the parameter inaccuracy like in the case of our work. Ryu et al. (2009) showed that because of the nonlinearity of soil water models the bias can appear even if an ensemble of model forecasts originates from Gaussian variations. Bias in surface soil moisture states can lead to significant mass balance errors and degrade the performance of the EnKF analysis in deeper soil layers. Overall, bias-blind data assimilation appears to result in biased and noisy updates (Dee, 2005). The data assimilation to correct simultaneously both parameters and state variables, that is, hydraulic conductivity and soil water contents, may be a way to develop a reliable soil water model for a specific site (e.g., Montaldo and Albertson, 2003). Systematic procedures for

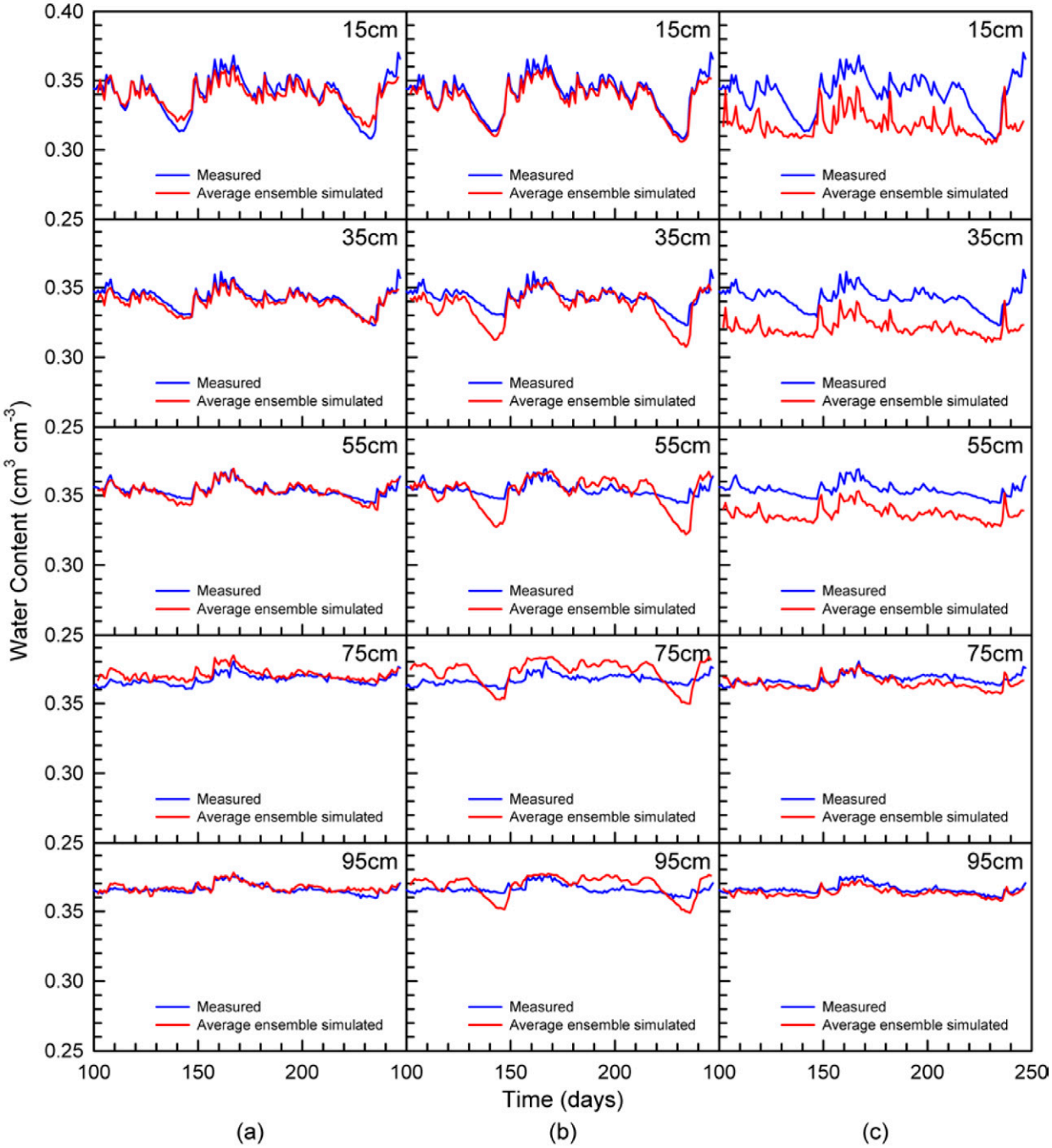

Fig. 6. Selected results of daily ensemble simulations update; (a) update with assimilation of data from sensors from all five depths, (b) update with assimilation of data from sensors at the 15-cm depth, and (c) update with assimilation of data from the sensors at the 95-cm depth.

such dual estimation in hydrologic models were introduced by Moradkhani et al. (2005a, 2005b) and Vrugt et al. (2005). Data assimilation to determine parameters of Richards' equation was recently demonstrated with a synthetic dataset (Montzka et al., 2011). The authors showed that, in the assimilation of surface water content data, the bias can be affected by the availability of information about water contents in the lower part of the profile and by soil properties. Approaches designed to model bias per se to improve data assimilation results for state variables have also been proposed (e.g., Dee, 2005). Evaluating the bias removal methods to apply with data assimilation from soil water content sensors in the Richards' equation model presents an interesting avenue for further research.

The soil water content DA with the EnKF provided good results across the whole soil profile even when the data from one depth or from two depths were used for the assimilation (Fig. 5 and 6 and Tables 3 and 4). This feature makes EnKF DA in soil water flow modeling more attractive as compared with direct insertion and other DA methods (Das and Mohanty, 2006). This is probably due to the fact that the assumptions of the soil water flow model have been applicable to the site conditions during the observation period which did not include events

Table 4. Root-mean-squared errors of soil water content simulations with calibrated models.

Pedotransfer function or ensemble	Sensor depth	RMSE (volume %) at the depth of				
		15 cm	35 cm	55 cm	75 cm	95 cm
		No data assimilation				
PTF 1	na†	1.03	0.93	1.16	0.82	0.57
PTF 2	na	1.22	1.04	0.84	1.04	0.58
PTF 3	na	1.89	0.79	0.82	1.19	0.58
PTF 4	na	1.11	0.96	0.9	0.89	0.63
PTF 5	na	1.47	1.34	1.54	1.67	1.36
PTF 6	na	1.4	1.51	1.08	1.03	0.84
		Weekly data assimilation				
Ensemble	All	0.74	0.8	0.7	0.64	0.37
Ensemble	15 cm	0.75	0.88	0.83	0.73	0.43
Ensemble	35 cm	0.81	0.76	0.71	0.72	0.45
Ensemble	55 cm	0.86	0.77	0.72	0.71	0.46
Ensemble	75 cm	0.77	0.86	0.82	0.7	0.43
Ensemble	95 cm	0.8	0.87	0.85	0.73	0.38
Ensemble	15,35 cm	0.69	0.87	0.83	0.79	0.49
Ensemble	15,55 cm	0.72	0.84	0.79	0.72	0.48
Ensemble	15,75 cm	0.74	0.86	0.82	0.68	0.38
Ensemble	15,95 cm	0.7	0.86	0.82	0.75	0.43
Ensemble	35,55 cm	0.9	0.75	0.71	0.72	0.47
Ensemble	35,75 cm	0.85	0.76	0.76	0.69	0.38
Ensemble	35,95 cm	0.84	0.78	0.78	0.76	0.43
Ensemble	55,75 cm	0.88	0.78	0.71	0.67	0.39
Ensemble	55,95 cm	0.86	0.81	0.72	0.68	0.35
Ensemble	75,95 cm	0.8	0.86	0.82	0.68	0.36

† na, not applicable.

conducive for preferential flow or long dry spells when Richards' equation might not be applicable. Satisfactory results across the whole soil profile were obtained from the assimilation of water contents at the top of the profile. This is in line with results and conclusion of works that have used the EnKF to assimilate the remotely sensed data at the soil surface to reproduce the water contents in soil profile (Das and Mohanty, 2006). If the physics is right then the model is able to capture the process if the boundary conditions are corrected. Crow and Wood (2003) noted that inadequacies in land surface model physics can create specific challenges in assimilation of soil surface water content data. However, the assimilated water content does not need to be measured on soil surface. In essence, a single soil water sensor or tensiometer can provide enough information to correct the performance of a physically-based soil hydraulic model for the whole soil profile. The top part of the profile was the preferable location of the sensors for assimilation in this work. However, the research site had no vegetation and soil water dynamics was not affected by shallow groundwater or by intensive evaporation. Should soil water dynamics be very different from the one in this work, a site-specific research would be needed to establish preferable locations of soil water content sensors.

Using the information about the temporal stability of soil water content measurements was essential for the data assimilation procedure. The "naive" standard errors

Table 5. Correlation coefficients between water contents measured at five depths after the removal of bias according the temporal stability model.

Time	Correlation coefficient									
	15 cm				35 cm			55 cm		75 cm
	35 cm	55 cm	75 cm	95 cm	55 cm	75 cm	95 cm	75 cm	95 cm	95 cm
days										
100	0.586	0.070	0.823	0.681	0.234	0.243	0.601	–0.085	0.534	0.441
107	0.705	–0.169	0.704	0.755	0.309	0.346	0.709	–0.260	0.191	0.634
114	0.516	–0.203	–0.050	0.526	–0.415	0.378	0.151	–0.332	0.179	0.210
121	0.762	–0.135	0.732	0.608	–0.280	0.782	0.775	–0.411	0.150	0.474
128	0.836	0.069	0.549	0.622	–0.009	0.661	0.632	–0.203	–0.066	0.772
135	0.552	–0.151	0.382	0.311	–0.185	0.199	0.099	–0.585	0.251	0.009
142	0.608	–0.141	0.255	0.055	0.149	0.059	–0.025	–0.331	0.569	0.091
149	0.122	–0.069	–0.092	–0.275	0.345	–0.169	0.198	0.197	0.253	–0.120
156	0.746	0.236	0.436	0.061	–0.008	0.588	0.377	–0.376	–0.030	0.132
163	0.197	0.191	–0.116	0.093	0.416	–0.133	0.018	–0.011	–0.060	0.512
170	0.192	0.284	–0.154	0.081	–0.022	0.400	0.270	–0.239	0.188	0.506
177	0.823	0.131	0.185	0.049	0.388	0.210	0.209	–0.210	0.308	0.365
184	0.539	0.160	–0.355	–0.456	0.097	0.035	0.033	–0.196	0.542	0.516
191	0.424	–0.215	0.280	–0.132	–0.028	0.209	0.175	0.067	0.406	–0.131
198	0.261	–0.069	0.087	0.288	0.492	0.243	–0.305	0.473	–0.218	–0.068
205	0.534	0.426	0.155	–0.082	0.738	0.237	0.265	0.081	0.263	0.382
212	0.723	0.210	–0.094	–0.602	–0.053	0.031	–0.206	0.052	–0.457	0.516
219	0.650	0.223	0.500	0.059	–0.150	0.591	–0.040	0.253	–0.029	0.289
226	0.666	0.096	0.591	0.046	–0.243	0.592	–0.013	0.161	–0.075	0.253
233	0.652	0.446	0.779	0.173	–0.110	0.588	–0.023	0.444	0.002	0.300
240	0.506	0.435	0.419	0.110	0.762	0.667	–0.247	0.631	0.105	0.173
247	–0.076	0.109	–0.011	0.288	0.540	0.605	0.616	0.405	0.508	0.637

of soil water contents at the five observation depths (15, 35, 55, 75, and 95 cm) were in the ranges 0.022 to 0.030, 0.017 to 0.023, 0.013 to 0.019, 0.019 to 0.023, and 0.020 to 0.028, respectively. The standard errors of the noise values in Eq. [2] for the same depths in the same order ranged from 0.004 to 0.011, 0.003 to 0.012, 0.003 to 0.009, 0.004 to 0.012, and 0.003 to 0.007. On average the standard errors of noise were about 30% of the naive standard errors. These results are similar to results of Starr (2005) who worked at the coarser scale and found that the temporal stability model explained 47% of the observed variability in soil water content whereas an additional 20% of the variability was attributed to random measurement error. Using errors of noise instead of naive errors in this work made the uncertainty in soil water content data about one order of magnitude smaller than the uncertainty in modeling results and caused the updated modeling results to be close to measurements. We realize that there was some underestimation of the data error in this work because the errors of the TDR measurement per se were not considered.

Decrease in the estimated measurement noise is achieved if the temporal stability model (Eq. [14]) is applicable. Appropriateness of this model depends on location and number of sensors. If sensor locations are such that the applicability temporal stability model is limited, the bias values may be inaccurate and noise values may be excessively large and updates will not lead to significant changes in simulated water contents. Sensor placement may also affect the value of the average water content that is used in updates with Kalman gain matrix (Eq. [7] and [8]). Although there were suggestions on the selection of locations where soil water content sensors should be placed to record values of water content close to the average over the study area (Grayson and Western, 1998; Jacobs et al., 2004), there are no general recommendations on selection of such sensor locations. The additional difficulty is that locations for sensors representing average over the study area may be different for different soil depths (Guber et al., 2008). Finding the environmental factors that may indicate probable locations of representative soil moisture measurements will help to decrease the number of sensors and improve the effectiveness of the soil water content data assimilation.

Several arbitrary choices were made in the design of this work. They included the decision to calibrate only saturated hydraulic conductivity rather than conductivity and water retention, limit the calibration period to 30 d, limit the number of water retention PTFs to six, and the number of K_{sat} values to four. We demonstrated that calibrating only saturated hydraulic conductivity values provides high accuracy of ensemble simulations. Calibrating van Genuchten water retention parameters could further improve accuracy of simulations with PTFs as it was shown for this dataset in the comprehensive calibration study (Jacques et al., 2002). However, calibrating 30 (four in Eq. [12] and two in Eq. [13] at five depths) requires using long time series to capture both long drying and extreme wetting events. The data assimilation needs to be applied just because accumulation of such exhaustive dataset takes time and may not be feasible with available resources. The number of calibrated parameters could be decreased by decreasing the number of hydrologically different layers, for example, setting this number equal to the number of soil genetic horizons as shown in Fig. 1. However, the differences in texture and organic carbon (OC) at different depths within horizons that we encountered would be ignored in such case. The number of models in ensemble has not been varied although it is known that the accuracy of assimilation results is affected by the increase of ensemble size (Houtekamer and Mitchell, 1998). We demonstrated that the satisfactory data assimilation from single depths is possible with the 24 models in the ensemble. However, this number has to be researched in specific applications. We have also not attempted to apply the quickly developing techniques of model calibration with data assimilation (Montzka et al., 2011) which represent a very promising avenue for soil hydrology research.

Data assimilation methods other than EnKF can also be applied to assimilate soil water contents measurements in soil water flow simulations. Sabater et al. (2007) compared several methods of DA for a soil–vegetation–atmosphere model with two soil layers and concluded that the EnKF was one of the best to use. However, it is not known how model specific such conclusions may be. The need in using other than EnKF DA methods may be caused by model-specific violations of the EnKF assumptions. Specifically, EnKF requires the normality of model and data errors distributions (Eq. [2] and [4]). The distributions of water contents simulated with the PTF generated model ensemble conformed to the normality hypothesis in the majority of cases. However, in some cases simulated soil water contents were not normally distributed. This percentage may be different in other soils and with other weather conditions. The DA methods, such as particle filtering, were proposed that do not require normality and the empirical distributions are generated from Monte Carlo simulations. These methods generally require the number of ensemble numbers much larger than EnKF (Weerts and El Serafy, 2005). Since the number of available PTFs is relatively small (Pachepsky and Rawls, 2004), a further research is needed to establish a procedure of creating large ensembles with relatively small numbers of PTFs.

CONCLUSIONS

Overall, this work demonstrated that bringing together developments in pedotransfer functions, temporal stability of soil water patterns, and soil water content sensors can create a new source of data to improve modeling results in soil hydrology and related fields. We observed that pedotransfer functions for saturated hydraulic conductivity in combination with the standard Mualem–van Genuchten model of unsaturated hydraulic conductivity created substantial bias in simulations of water contents in soil profile. Caution has to be exercised in using K_{sat} PTFs, and they may need further development to be used in applications at the pedon or the field scale. Assimilation of soil water content sensor data appeared to be very effective

in correcting soil water content profiles simulated with the Richards' equation based model; small number of sensors was sufficient to correct the simulated profile. The efficiency of assimilation increased with the frequency of updates.

APPENDIX

Six PTFs were used to estimate the water retention parameters in this study to develop the ensemble of models. The PTFs with input soil properties are listed in Table A.1. Two PTFs developed by Wösten et al. (1999) estimate the parameters of van Genuchten equation (Eq. [12]), and another four PTFs derive the van Genuchten parameters by fitting Eq. [12] to the estimated water contents at selected capillary pressures (Tomasella and Hodnett, 1998; Gupta and Larson, 1979; Rajkai and Varallyay, 1992; Rawls et al., 1983). The details of the six pedotransfer equations are listed below. The PTFs were used with textural composition shown in Fig. 1, OC contents of 2.2, 0.8, 0.4, 0.3, and 0.6% at depths of 15, 35, 55, 75, and 95 cm, respectively, and bulk density by soil horizons Ap, C1, and C2 of 1.42, 1.54, and 1.53 g cm^{-3}, respectively.

Wösten et al. (1999) derived class PTFs based on the all-Europe database HYPRES and the van Genuchten parameters were obtained by fitting Eq. [12] to geometric mean water contents for five textural groups (Table 2 of Guber and Pachepsky, 2010). Wösten et al. (1999) also derived regression equations to estimate the van Genuchten parameters from soil texture, OC, and soil bulk density (ρ_b) (Eq. [27], [28], [29] in Guber and Pachepsky, 2010; not included here because of their size).

Tomasella and Hodnett (1998) derived regression parameters for water content (θ) at nine values of soil matric potential based on the nationwide Brazilian soil database:

$$\theta = 0.01(a \times \mathrm{OC} + b \times \mathrm{silt} + c \times \mathrm{clay} + d) \quad [A.1]$$

where a, b, c, and d are regression coefficients listed in Table 3 of (Guber and Pachepsky, 2010).

Rawls et al. (1983) developed 12 regression equations to relate the soil water contents at 12 capillary pressures to sand, clay, OC contents, and bulk density using the U.S. Cooperative Soil Survey Database from Rawls et al. (1982):

$$\theta = a + b \times \mathrm{sand} + c \times \mathrm{clay} + d \times \mathrm{OC} + e \times \rho_b \quad [A.2]$$

where a, b, c, d, and e are coefficients of the linear regression equations listed in Table 5 of Guber and Pachepsky (2010).

Gupta and Larson (1979) derived predictive equations for the water content at 12 capillary pressures using a subset of the U.S. National Cooperative Survey database:

$$\theta = a \times \mathrm{sand} + b \times \mathrm{silt} + c \times \mathrm{clay} + d \times \mathrm{OC} + e \times \rho_b \quad [A.3]$$

where a, b, c, d, and e are coefficients of the linear regression equations to predict soil water content at specific capillary pressure listed in Table 6 of Guber and Pachepsky (2010).

Rajkai and Varallyay (1992) developed a nonlinear regression equation for 10 matric potential levels using a Hungarian nationwide database:

$$\theta = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 + b_4 X_1^2 + b_5 X_2^2 \quad [A.4]$$

where b_0, b_1, b_2, b_3, b_4, b_5, and X_1, X_2 are coefficients and variables of the nonlinear regression equations, respectively. The coefficients and variables at eight capillary pressures used in this study are listed in Table 7 of Guber and Pachepsky (2010).

Table A.1. List of soil water retention pedotransfer functions (PTFs) and estimated parameters α and n in the van Genuchten equation.

PTF	Depth	Wösten et al. (1999)		Tomasella and Hodnett (1998)	Gupta and Larson (1979)	Rajkai and Varallyay (1992)	Rawls et al. (1983)
Model		VG†	VG†	WH→VG‡	WH→VG‡	WH→VG‡	WH→VG‡
Clay, %		+			+	+	+
Silt, %		+	+	+	+	+	+
Sand, %		+	+	+	+	+	+
Organic C, %			+	+	+		+
Bulk density, g cm^{-3}			+		+	+	+
α (1/m)	15 cm	0.0249	0.0436	0.1705	0.0281	0.0084	0.0532
	35 cm	0.0314	0.0404	0.1118	0.0405	0.0064	0.0477
	55 cm	0.0314	0.045	0.1034	0.0488	0.0061	0.0527
	75 cm	0.0314	0.0394	0.0734	0.0353	0.0059	0.0394
	95 cm	0.0314	0.028	0.0515	0.0192	0.0062	0.03
n	15 cm	1.1689	1.2214	1.2097	1.4158	1.1827	1.2916
	15 cm	1.1804	1.2537	1.2173	1.3566	1.1672	1.3455
	55 cm	1.1804	1.2593	1.2339	1.3385	1.182	1.3567
	75 cm	1.1804	1.2376	1.2318	1.3188	1.1214	1.3474
	95 cm	1.1804	1.2548	1.2259	1.368	1.0737	1.3539

† Parameters of the van Genuchten equation are estimated with the pedotransfer function

‡ Water contents at specific pressure heads are estimated, and then the van Genuchten equation is fitted to the estimates.

The K_{sat} values were estimated (a) based on textural class and bulk density according to the table developed from a large U.S. nationwide database (Rawls et al., 1998), (b) as the average values of K_{sat} found from three large databases (Schaap and Leij, 1998), (c) from clay and sand contents with regression equations developed from a large dataset of Soil Conservation Service (SCS) Soil Survey Information Reports (Carsel and Parrish, 1988), and (d) from fitting the van Genuchten–Mualem equation to geometric mean water contents developed using the European continental database HYPRES (Wösten et al., 1999). Estimation results are shown in Table A.2.

ACKNOWLEDGMENTS

This work was partially supported by the Interagency Agreement between U.S. Nuclear Regulatory Commission and USDA Agricultural Research Service.

Table A.2. List of saturated hydraulic conductivity from literature pedotransfer functions.

No.	Reference	Saturated hydraulic conductivity	
		Sandy loam	Loam
		– cm d^{-1} –	
1	Rawls et al. (1998)	55.0	12.5
2	Schaap and Leij (1998)	38.0	12.0
3	Carsel and Parrish (1988)	106.0	25.0
4	Wösten et al. (1999)	12.1	10.8

REFERENCES

Aboitiz, M., J.W. Labadie, and D.F. Heermann. 1986. Stochastic sol moisture estimation and forecasting for irrigated fields. Water Resour. Res. 22:180–190. doi:10.1029/WR022i002p00180

Carsel, R.F., and R.S. Parrish. 1988. Developing joint probability distributions of soil water retention characteristics. Water Resour. Res. 24:755–769. doi:10.1029/WR024i005p00755

Crow, W.T., and E. Van Loon. 2006. The impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J. Hydrometeorol. 8:421–431. doi:10.1175/JHM499.1

Crow, W.T., and E.F. Wood. 2003. The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. Adv. Water Resour. 26:137–149. doi:10.1016/S0309-1708(02)00088-X

Das, N.N., and B.P. Mohanty. 2006. Root zone soil moisture assessment using remote sensing and vadose zone modeling. Vadose Zone J. 5:296–307. doi:10.2136/vzj2005.0033

Dee, D.P. 2005. Bias and data assimilation. Q. J. R. Meteorol. Soc. C 131:3323–3343. doi:10.1256/qj.05.137

Entekhabi, D., H. Nakamura, and E.G. Njoku. 1994. Solving the inverse-problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. IEEE Geosci. Remote Sens. 32:438–448. doi:10.1109/36.295058

Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99:10143–10162. doi:10.1029/94JC00572

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

Fares, A., H. Hamdhani, V. Polyakov, A. Dogan, and H. Valenzuela. 2006. Real-time soil water monitoring for optimum water management. J. Am. Water Resour. Assoc. 42:1527–1535. doi:10.1111/j.1752-1688.2006.tb06018.x

Grayson, R.B., and A.W. Western. 1998. Towards areal estimation of soil water content from point measurements: Time and space stability of mean response. J. Hydrol. 207:68–82. doi:10.1016/S0022-1694(98)00096-1

Guber, A.K., T.J. Gish, Y.A. Pachepsky, M.Th. van Genuchten, C.S. Daughtry, T.J. Nicholson, and R.E. Cady. 2008. Temporal stability of soil water content and soil water flux across agricultural fields. Catena 73:125–133. doi:10.1016/j.catena.2007.09.010

Guber, A.K., and Y.A. Pachepsky. 2010. Multimodeling with pedotransfer functions. Documentation and user manual for PTF calculator (CalcPTF), Version 2.0. Beltsville Agricultural Research Center, USDA-ARS. http://www.ars.usda.gov/SP2UserFiles/ad_hoc/12655300EnvironmentalTransport/CalcPTFFiles/PTF_Manual.version_2.0.pdf (accessed 24 Feb. 2012).

Guber, A.K., Y.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. doi:10.2136/vzj2005.0111

Guber, A.K., Y.A. Pachepsky, M.Th. van Genuchten, J. Simunek, D. Jacques, A. Nemes, T.J. Nicholson, and R.E. Cady. 2009. Multimodel simulation of water flow in a field soil using pedotransfer functions. Vadose Zone J. 8:1–10. doi:10.2136/vzj2007.0144

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. doi:10.1029/WR015i006p01633

Hamill, T.M., J.S. Whitaker, and X. Wei. 2004. Ensemble re-forecasting: Improving medium-range forecast skill using retrospective forecasts. Mon. Weather Rev. 132:1434–1447. doi:10.1175/1520-0493(2004)1322.0.CO;2

Heathman, G.C., P.J. Starks, L.R. Ahuja, and T.J. Jackson. 2003. Assimilation of surface soil moisture to estimate profile soil water content. J. Hydrol. 279:1–17. doi:10.1016/S0022-1694(03)00088-X

Houser, P.R., W.J. Shuttleworth, J.S. Famiglietti, H.V. Gupta, K.H. Syed, and D.C. Goodrich. 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour. Res. 34:3405–3420. doi:10.1029/1998WR900001

Houtekamer, P.L., and H.L. Mitchell. 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. 126:796–811. doi:10.1175/1520-0493(1998)1262.0.CO;2

Huang, C., X. Li, L. Lu, and J. Gu. 2008. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens. Environ. 112:888–900. doi:10.1016/j.rse.2007.06.026

Jacobs, J.M., B.P. Mohanty, E.C. Hsu, and D. Miller. 2004. SMEX02: Field scale variability, time stability and similarity of soil water. Remote Sens. Environ. 92:436–446. doi:10.1016/j.rse.2004.02.017

Jacques, D. 2000. Analysis of water flow and solute transport at the field scale. Ph.D. thesis no. 454. Faculteit Landbouwkundige en Toegepaste Wetenschappen, K.U. Leuven, Belgium.

Jacques, D., B. Mohanty, A. Timmerman, and J. Feyen. 2001. Study of time dependency of factors affecting the spatial distribution of soil water content in a field-plot. Phys. Chem. Earth (B) 26:629–634. doi:10.1016/S1464-1909(01)00060-0

Jacques, D., J. Simunek, A. Timmerman, and J. Feyen. 2002. Calibration of Richards' and convection-dispersion equations to field-scale water flow and solute transport under rainfall conditions. J. Hydrol. 259:15–31. doi:10.1016/S0022-1694(01)00591-1

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82:35–45. doi:10.1115/1.3662552

Kelleners, T.J., R.W.O. Soppe, J.E. Ayars, J. Šimůnek, and T.H. Skaggs. 2005. Inverse analysis of upward water flow in a groundwater table lysimeter. Vadose Zone J. 4:558–572. doi:10.2136/vzj2004.0118

Lahoz, W., B. Khattatov, and R. Ménard, editors. 2010. Data assimilation: Making sense of observations. Springer, New York.

Mallants, D., D. Jacques, P. Tseng, M.Th. van Genuchten, and J. Feyen. 1997. Comparison of hydraulic measurements using three sizes of soil cores. J. Hydrol. 199:295–318. doi:10.1016/S0022-1694(96)03331-8

Mandell, J. 2007. A brief tutorial on the ensemble Kalman filter. Cornell Univ. http://arxiv.org/PS_cache/arxiv/pdf/0901/0901.3725v1.pdf (accessed 24 Feb. 2012).

Montaldo, N., and J.D. Albertson. 2003. Multi-scale assimilation of surface soil moisture data for robust root zone moisture prediction. Adv. Water Resour. 26:33–44. doi:10.1016/S0309-1708(02)00103-3

Montzka, C., H. Moradkhani, L. Weihermüller, H.-J. Hendricks Franssen, M. Canty, and H. Vereecken. 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399:410–421. doi:10.1016/j.jhydrol.2011.01.020

Moradkhani, H., K. Hsu, H.V. Gupta, and S. Sorooshian. 2005b. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using particle filter. Water Resour. Res. 41:W05012 10.1029/2004WR003604. doi:10.1029/2004WR003604

Moradkhani, H., S. Sorooshian, H.V. Gupta, and P. Houser. 2005a. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv. Water Resour. 28:135–147. doi:10.1016/j.advwatres.2004.09.002

Ng, G.-H.C., D. McLaughlin, D. Entekhabi, and B. Scanlon. 2009. Using data assimilation to identify diffuse recharge mechanisms from chemical and physical data in the unsaturated zone. Water Resour. Res. 45:W09409 10.1029/2009WR007831. doi:10.1029/2009WR007831

Or, D., and R.J. Hanks. 1992. Spatial and temporal soil water estimation considering soil variability and evapotranspiration uncertainty. Water Resour. Res. 28:803–814. doi:10.1029/91WR02585

Pachepsky, Y.A., A.K. Guber, and D. Jacques. 2005. Temporal persistence in vertical distributions of soil moisture contents. Soil Sci. Soc. Am. J. 69:347–352. doi:10.2136/sssaj2005.0347

Pachepsky, Y.A., and W.J. Rawls, editors. 2004. Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam.

Quinn, G.P., and M.J. Keough, editors. 2002. Experimental design and data analysis for biologists. Cambridge Univ. Press, New York.

Rajkai, K., and G. Varallyay. 1992. Estimating soil water retention from simpler properties by regression techniques. In: M.Th. van Genuchten et al.,

editors, Proceedings International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. Univ. of California, Riverside. p. 417–416.

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.L. Brakensiek, and B. Soni. 1983. Agricultural management effects on soil water processes. Part I. Soil water retention and Green-Ampt parameters. Trans. ASAE 26:1747–1752.

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.

Ryu, D., W.T. Crow, X. Zhan, and T.J. Jackson. 2009. Correcting unintended perturbation biases in hydrologic data assimilation. J. Hydrometeorol. 10:734–750. doi:10.1175/2008JHM1038.1

Sabater, J.M., L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. de Rosnay. 2007. From near-surface to root-zone soil moisture using different assimilation techniques. J. Hydrometeorol. 8:194–206. doi:10.1175/JHM571.1

Santanello, J.A., Jr., C.D. Peters-Lidard, M.E. Garcia, D.M. Mocko, M.A. Tischler, M.S. Moran, and D.P. Thoma. 2007. Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed. Remote Sens. Environ. 110:79–97. doi:10.1016/j.rse.2007.02.007

Schaap, M.G., and F.J. Leij. 1998. Database-realated accuracy and uncertainty of pedotransfer functions. Soil Sci. 163:765–779. doi:10.1097/00010694-199810000-00001

Schaap, M.G., and F.J. Leij. 2000. Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model. Soil Sci. Soc. Am. J. 64:843–851. doi:10.2136/sssaj2000.643843x

Schaffer, B. 1998. Flooding responses and water-use efficiency of subtropical and tropical fruit trees in an environmentally sensitive wetland. Ann. Bot. (London) 81:475–481. doi:10.1006/anbo.1998.0593

Šimůnek, J., M. Šejna, H. Saito, M. Sakai, and M.Th. van Genuchten. 2008. The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Version 4.0. Univ. of California, Riverside.

Starr, G.C. 2005. Assessing temporal stability and spatial variability of soil water patterns with implications for precision water management. Agric. Water Manage. 72:223–243. doi:10.1016/j.agwat.2004.09.020

Tomasella, J., and M.G. Hodnett. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Sci. 163:190–202. doi:10.1097/00010694-199803000-00003

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. doi:10.2136/sssaj1980.03615995004400050002x

Vereecken, H., J.A. Huisman, H. Bogena, J. Vanderborght, J.A. Vrugt, and J.W. Hopmans. 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resour. Res. 44:W00D06 10.1029/2008WR006829. doi:10.1029/2008WR006829

Vrugt, J.A., C.G.H. Diks, H.V. Gupta, W. Bouten, and J.M. Verstraten. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour. Res. 41:W01017 10.1029/2004WR003059. doi:10.1029/2004WR003059

Walker, J.P., G.R. Willgoose, and J.D. Kalma. 2001a. One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. J. Hydrometeorol. 2:356–373. doi:10.1175/1525-7541(2001)0022.0.CO;2

Walker, J.P., G.R. Willgoose, and J.D. Kalma. 2001b. One-dimensional soil moisture profile retrieval by assimilation of near surface observations: A comparison of retrieval algorithms. Adv. Water Resour. 24:631–650. doi:10.1016/S0309-1708(00)00043-9

Weerts, A.H., and G.Y.H. El Serafy. 2005. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res. 42:W09403 10.1029/2005WR004093. doi:10.1029/2005WR004093

Wendroth, O., H. Rogasik, S. Koszinski, C.J. Ritsema, L.W. Dekker, and D.R. Nielsen. 1999. State-space prediction of field-scale soil water content time series in a sandy loam. Soil Tillage Res. 50:85–93. doi:10.1016/S0167-1987(98)00201-3

Wigley, T.M.L., K.R. Briffa, and P.D. Jones. 1984. On the average value of correlated time series with applications in Dendroclimatology and Hydrometeorology. J. Clim. Appl. Meteorol. 23:201–213. doi:10.1175/1520-0450(1984)0232.0.CO;2

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. doi:10.1016/S0016-7061(98)00132-3