Information Content of Data for Identifying Soil Hydraulic Parameters from Outflow Experiments

J. A. Vrugt,* W. Bouten, and A. H. Weerts

ABSTRACT

Process-oriented models of open systems often contain parameters that cannot be measured directly but can only be obtained by inverse modeling. A conventional inverse method is typically based on the minimization of an objective function that lumps the discrepancies in time series of observed values and predicted model response. However, problems are often encountered with the non-uniqueness of the parameter estimates. Non-unique parameter estimates result in case of low parameter sensitivity, mutual parameter dependency, and high measurement noise. These problems can be solved partly if we do not use the entire data set but focus on subsets where the model is most sensitive to changes in the unknown parameters. Therefore we propose PIMLI (Parameter Identification Method based on the Localization of Information), that uses the variability in time of the model sensitivity for the various parameters to split the total set of measurements into disjunctive subsets that each contain the most information on one of the model parameters. Thereupon, each distinguished subset is used to constrain its corresponding parameter. To illustrate PIMLI we chose a simulated multi-step outflow (MSO) experiment in which only cumulative outflow is measured because of its well-known problems with the uniqueness of the identified soil hydraulic properties. The results show that PIMLI not only leads to unique parameter sets of soil hydraulic properties for a range of soils but also significantly improves the understanding of uniqueness problems related to parameter identification.

Models for environmental applications vary in sophistication and complexity, ranging from simple data-oriented models to highly complex process-oriented models. These models give an approximate description of the system under study and contain several unknown quantities, such as parameters. Often, these model parameters cannot be measured directly but can only be obtained by inverse modeling. If these models are to be applied without calibration, then transfer functions must be found to link these model parameters to other properties that can be measured independently (Schaap et al., 1998). The uniqueness of a set of calibrated parameters is a prerequisite to finding these transfer functions.

The simplest form of parameter estimation is curve fitting in which measured data are represented by a static function with parameters that provide the best possible fit to the data (van Genuchten et al., 1991). A more complex form of parameter estimation is inverse modeling, where parameters are optimized while minimizing a suitable objective function that expresses the discrepancy between the output of a dynamic model and the measurements (Janssen and Heuberger, 1995). However, using this conventional approach, problems are often encountered with the non-uniqueness of the optimized parameters. Non-uniqueness leads to more than one set of parameters, each yielding minimum values for the objective function determined by local minima or by the same global minimum at more than one point in the parameter space (Gupta and Sorooshian, 1985; Duan et al., 1992).

Seemingly there was a widespread conviction that the best way to solve the non-uniqueness problem was to include additional and more accurate measurements (Eching and Hopmans, 1993; Van Dam et al., 1994). However, research into data requirements has led to the understanding that the information content of the data is far more important than the amount of data used for model calibration (Kuczera, 1982; Sorooshian et al., 1983; Gupta and Sorooshian, 1985; Yapo et al., 1996; Gupta et al., 1998). Therefore, we developed a parameter identification technique entitled PIMLI (Parameter Identification Method based on the Localization of Information) that treats the data in such a way as to preserve the specific information with respect to the various parameters and that uses the information content of data to identify the model parameters. For illustrating PIMLI we used the MSO experiment that only measures cumulative outflow, because of its well-known problems regarding the non-uniqueness of the identified parameters for unsaturated flow (Hopmans and Šimůnek, 1999).

The use of inverse methods for determining the unsaturated flow parameters from transient experiments was first reported by Zachmann (1981). Kool et al. (1985a,b) used this inverse modeling approach to determine soil hydraulic properties from one-step outflow (OSO) experiments, but experienced problems with the non-uniqueness of the parameter estimation. Further investigations of the inverse method demonstrated the need for additional $\theta(h)$ data (Hudson et al., 1991; Van Dam et al., 1992; Bohne et al., 1993) or tensiometer measurements inside the soil sample (Kool and Parker, 1988; Toorman et al., 1992; Eching and Hopmans, 1993) to overcome the problem of non-uniqueness of parameters. The benefit of including tensiometer measurements is apparent, as the optimized soil water retention is forced to match the observed $\theta(h)$ data. Another way to realize more reliable parameter estimates is to incrementally increase the pneumatic pressure in several steps, the MSO experiment, instead of a single pressure increment (Van Dam et al., 1994).

The objective of this study was to analyze the tempo-

J.A. Vrugt, W. Bouten, and A.H. Weerts, Institute for Biodiversity and Ecosystem dynamics, University of Amsterdam, Nieuwe Prinsengracht 130, Amsterdam, 1018 VZ, the Netherlands. Received 17 Sept. 1999. *Corresponding author (j.vrugt@frw.uva.nl).

Published in Soil Sci. Soc. Am. J. 65:19–27 (2001).

Abbreviations: MSO, multi-step outflow; MVG, Mualem–van Genuchten; OSO, one-step outflow; PIMLI, Parameter Identification Method based on the Localization of Information; SWIF, soil water in forested ecosystems model.

ral variability of model sensitivity for the various soil hydraulic parameters for a MSO experiment that only measures cumulative outflow. These results were used to assess transient flow experiment potentials and limitations for parameter identification. For both purposes we used the PIMLI algorithm.

First, we evaluated the uniqueness of the inverse problem for a sandy soil using two-dimensional response surfaces of the objective function. These response surfaces were obtained by perturbing two selected soil hydraulic parameters around their true values, while maintaining the additional parameters constant at their true values (Toorman et al., 1992; Šimůnek et al., 1998). Then, we illustrated the PIMLI for the same sandy soil, a silty soil, and a clayey soil. Our study was carried out with numerically generated error-free data. These artificial measurements are preferred because the soil hydraulic parameters are then known beforehand (Toorman et al., 1992; Šimůnek et al., 1998).

MATERIALS AND METHODS

Water Flow Theory

The soil water in forested ecosystems (SWIF) model is used to simulate the MSO experiment. A full description is given by Tiktak and Bouten (1992). Transient flow is simulated by numerically solving the Richards equation (Eq. [1]) using the mass-conservative scheme proposed by Celia et al. (1990)

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

where h denotes soil water pressure head (L), θ is the volumetric water content ($L^3\ L^{-3}$), t is time (T), x is the spatial coordinate (L) (positive upward), and $K(h)$ is the unsaturated hydraulic conductivity ($L\ T^{-1}$). Because the user can control the upper and lower boundary in SWIF, it is possible to apply a zero flux upper boundary and suction at the lower boundary (Dirichlet condition), as required in the outflow experiment.

The soil hydraulic functions in SWIF are described by the Mualem–van Genuchten (MVG) model (Mualem, 1976; van Genuchten, 1980):

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

$$K(\theta) = K_s S_e^{\lambda}\left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2 \quad h < 0 \quad [3]$$

$$K(\theta) = K_s \quad h \geq 0 \quad [4]$$

where θ ($L^3\ L^{-3}$) denotes water content, θ_s is the saturated water content ($L^3\ L^{-3}$), θ_r is the residual water content ($L^3\ L^{-3}$), K_s is the saturated conductivity ($L\ T^{-1}$), and α (L^{-1}), n ($m = 1 - 1/n$), and λ are curve shape parameters.

Artificial Measurements

The artificial outflow measurements were calculated for a soil core with a diameter of 5.04 cm and a height of 5 cm, corresponding with a soil volume of 100 cm^3. At the bottom of the soil core, we simulated a 0.7-cm-thick porous ceramic plate with a saturated conductivity of 0.0034 m d^{-1}. Because the ceramic plate was considered part of the porous system in the numerical simulations, parameter values were chosen such that the ceramic remained fully saturated for the range of pressure head steps. As the initial condition, hydraulic equilibrium was assumed with a pressure head $h = -1.0$ cm at the bottom of the soil core, and $h = -6.0$ cm at the top. The following pressure head steps and time periods (in parentheses) were applied in the experiment: $h = -0.0030$ m ($0 \leq 0.5$ d), $h = -0.15$ m ($0.5 \leq 1.5$ d), $h = -0.50$ m ($1.5 \leq 3.5$ d), $h = -1.00$ m ($3.5 \leq 5.5$), $h = -3.00$ m ($5.5 \leq 12.5$ d), and $h = -5.00$ m ($12.5 \leq 20.0$ d).

Response Surface Analysis

The commonly used objective function OF(b), which is normally minimized with the help of a classical parameter optimization algorithm, is defined as

$$\mathrm{OF}(\mathbf{b}) = \sum_{j=1}^{m}\left\{w_j \sum_{i=1}^{n_j} w_{i,j}[q_j^*(t_i) - q_j(t_i,\mathbf{b})]^2\right\} \quad [5]$$

where $\mathbf{b}$ is the array of parameter values (θ_s, θ_r, α, n, K_s, and λ), j represents the different sets of measurements (cumulative outflow, water content, and flux density), n_j is the number of measurements within a particular set, $q_j^*(t_i)$ are measurements of type j at time t_i, $q_i(t_i,\mathbf{b})$ are the corresponding model predictions using the parameters in $\mathbf{b}$, and w_j and $w_{i,j}$ are weighting factors associated with measurement type and individual measurements, respectively. Because all the measurements were generated with the numerical model, we assumed that the errors associated with individual measurements of all types were identical. Therefore, $w_{i,j}$ was set equal to 1 for all the measurements. Differences in weighting between measurement types, as caused by differences in magnitude and their number n_j, were normalized by dividing each data point by both the variance of the measurements of type j and the n_j (Clausnitzer and Hopmans, 1995)

$$w_j = \frac{1}{n_j}\left(\frac{1}{\sigma_j^2}\right) \quad [6]$$

where σ_j and n_j denote the standard deviation and the number of j-type measurements, respectively.

We investigated the posedness of the conventional inverse solution using two-dimensional response surfaces of the objective function. We approached this question in a way similar to that done previously by Toorman et al. (1992) and Šimůnek et al. (1998). The response surfaces are obtained by solving the objective function in Eq. [5] for many possible combinations of the selected pair of parameter values, while keeping the additional four parameters constant at their true values. These response surfaces reveal the occurrence of local minima, the presence of a well-defined global minimum, the parameter sensitivity, and parameter correlations. If response surfaces do not display a well-defined global minimum in the two-dimensional parameter planes, the conventional inverse parameter estimation technique may certainly be expected to be unsuccessful in a multidimensional plane. The response surfaces were calculated on a rectangular grid with parameter values corresponding with the sandy soil in Table 1. Each parameter domain

Table 1. Soil hydraulic parameters of the sandy soil and parameter ranges used for the parameter planes and the Latin Hypercube sampling strategy.

Parameter	Sandy soil†	Min.	Max.	Unit	SD
θ_s	0.38	0.20	0.50	$m^3\ m^{-3}$	0.0866
θ_r	0.02	0.00	0.07	$m^3\ m^{-3}$	0.0202
α	2.14	0.10	3.00	m^{-1}	0.8374
n	2.075	1.05	3.00	–	0.5631
K_s	0.1556	1.10^{-4}	0.50	m d^{-1}	0.1443
λ	0.039	−2.00	2.00	–	1.1550

† Adapted from Wösten et al. (1994).

was discretized into 40 equidistant discrete points with domain ranges as presented in Table 1, resulting in 1600 grid points for each response surface.

By limiting the number of parameters within a single response surface analysis, the behavior of the objective function in the different parameter planes can only suggest how the objective function might behave in the full parameter space. For example, local minima of the objective function could exist, but not appear in the cross-sectional planes (Šimůnek and van Genuchten, 1996). Nevertheless, response surfaces are a suitable tool to obtain insight into the behavior of the objective function in the full parameter space.

PIMLI

PIMLI is a parameter identification method that for each parameter uses a separate subset of measurements with the highest information content for that particular parameter. If we want to identify those different subsets, we need a stochastic approach (Hollenbeck and Jensen, 1998) since analytical solutions for transient water flow are not available and cannot be derived. Therefore, the numerical model SWIF was first used to generate 1000 Monte-Carlo simulations of a MSO experiment with given initial and boundary conditions by randomly selecting combinations of the MVG parameters θ_s, θ_r, α, n, K_s, and λ. The parameter sets were selected with a Latin-Hypercube sampling method (McKay et al., 1979) in which parameter ranges in Table 1 were used. We assumed no correlation between the different soil hydraulic parameters. The outcome of the Monte Carlo analysis resulted in a population of generated model outputs at every discrete time value. Thereafter, a reference run of measurements was simulated with SWIF using hydraulic properties corresponding with the sandy soil (Table 1). The cumulative outflow, its first derivative (flux density), and the average water content of this reference run as deduced from the cumulative outflow and the water content at the end of the experiment are presented in Fig. 1A.

Each single artificial measurement of the sandy soil was then used to select those individuals of the total Monte Carlo population, which fit that particular measurement within its accuracy interval (Musters and Bouten, 2000). Since all measurements are subjected to experimental errors, PIMLI includes this data error in the analysis. The errors of the different measurements were taken to be

$$\Delta_Q = 0.05$$

$$\Delta_\theta = \frac{\Delta_Q}{100} + \Delta_{\theta,\mathrm{end}}$$

$$\Delta_F = \frac{2\Delta_Q}{100\pi r^2 \Delta t} \qquad [7]$$

where Δ_Q (cm^3), Δ_θ (cm^3 cm^{-3}), and Δ_F (m d^{-1}) denote the error of cumulative outflow, water content, and flux density, respectively; $\Delta_{\theta,\mathrm{end}}$ is the error in water content of the soil sample as caused by weighing at the end of the experiment, r denotes the radius of the soil core, and Δt is the time interval between two subsequent measurements, which was chosen as 0.05 d. The error of the outflow was determined from the

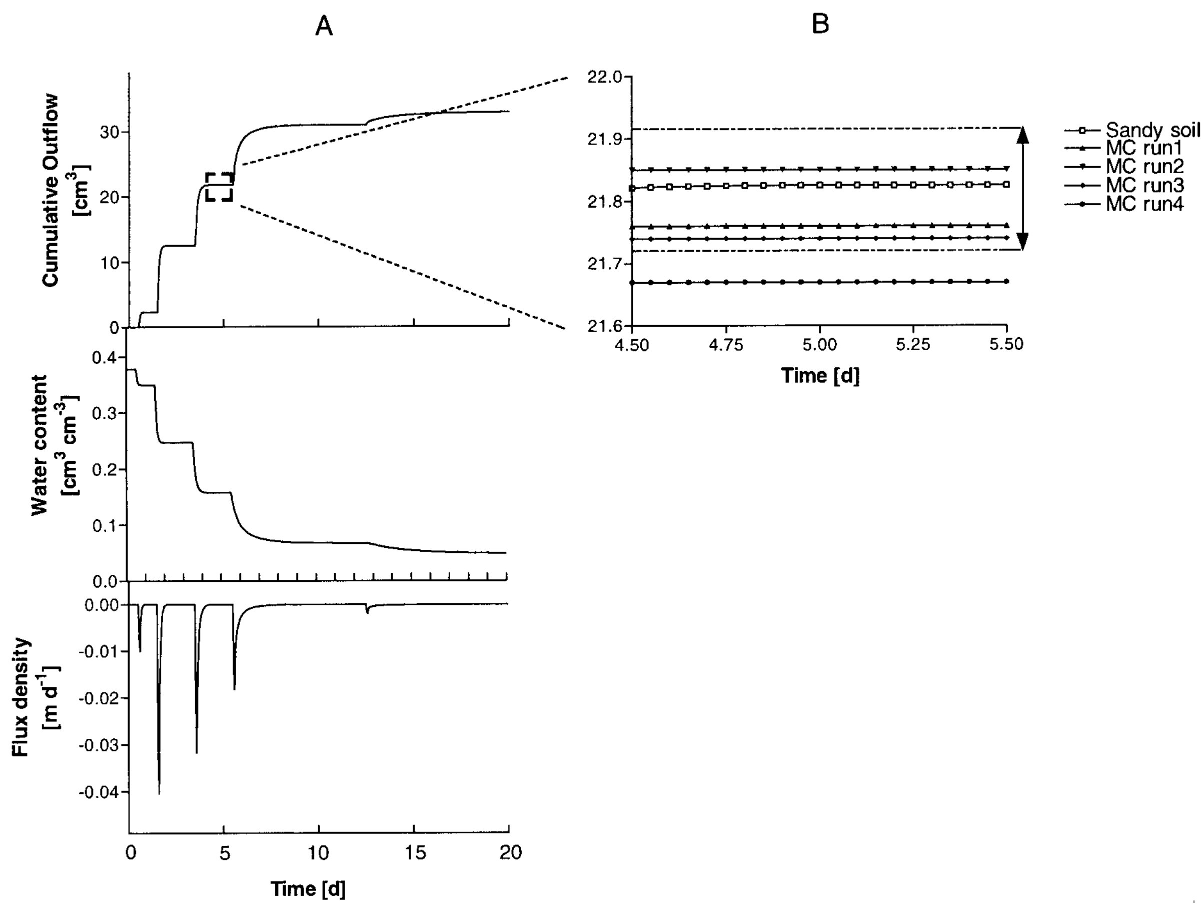

Fig. 1. Simulated cumulative outflow, water content, and flux density of the sandy soil as function of time during the multi-step outflow experiment. MC refers to Monte Carlo simulation.

resolution and accuracy of pressure transducers that are used for automated monitoring of the outflow dynamics. We assumed $\Delta_{\theta,\mathrm{end}}$ to be 0.01 ($\mathrm{cm^3\ cm^{-3}}$).

A parameter set was accepted if its corresponding simulation fits a particular measurement of the reference run within an accuracy range of either 0.1 $\mathrm{cm^3}$ ($2\Delta_Q$) for cumulative outflow 0.021 $\mathrm{cm^3\ cm^{-3}}$ for water content ($2\Delta_\theta$) or 2.10^{-3} m $\mathrm{d^{-1}}$ in terms of flux density ($2\Delta_F$). For example, in Fig. 1B at t = 5.0 d, the cumulative outflow of the reference run equals 21.82 $\mathrm{cm^3}$, so all Monte Carlo simulations are accepted that have a cumulative outflow at t = 5.0 that lie between 21.72 and 21.92 $\mathrm{cm^3}$. Hence, at t = 5.0, the Monte Carlo Runs 1, 2, and 3 are accepted and Monte Carlo Run 4 is rejected. At every point in time the measurement of the sandy soil is represented by a number of Monte Carlo simulations, with known soil hydraulic parameters. These vectors of accepted parameters are then used to calculate the standard deviation of each parameter.

PIMLI uses the following measure to quantify the sensitivity of a specific measurement for one of the model parameters

$$\mathrm{IC}_m(p) = 1 - \frac{\sigma(p)_m}{\sigma(p)} \qquad [8]$$

where $\mathrm{IC}_m(p)$ denotes the information content of measurement, m, with respect to the parameter, p, $\sigma(p)_m$ is the standard deviation of parameter, p, in the accepted sets at measurement, m, and $\sigma(p)$ is the standard deviation of the parameter, p, in the initial ranges. If $\mathrm{IC}_m(p)$ is close to zero, this implies that this measurement is hardly sensitive for parameter, p, whereas an information content close to one indicates that the information content or sensitivity of measurement, m, for parameter, p, is very high.

The time series of information content is then used to split the total time series of measurements into disjunctive subsets, with only measurements of high information content for a specific parameter. Hence, robust information for the least sensitive parameters will only appear if the most dominant parameters are close to their true value. Therefore, we followed an iterative procedure in which each iteration is used to identify a subset of measurements with highest information content for a not yet constrained parameter. Once these subsets are identified, then the histogram of the accepted parameters in this set is used for sampling in the next iteration. For this, the histogram is divided into 10 different equidistant classes in which the frequency of every class is used in a Latin Hypercube sampling.

To reduce the number of Monte Carlo simulations and still guarantee that enough simulations are accepted to ensure a representative value of the information content of a particular measurement, we temporarily adjusted Δ_Q to $5\Delta_Q$ for the localization of the different subsets after the first 1000 Monte Carlo simulations. Every next iteration, Δ_Q was reduced with 0.05 $\mathrm{cm^3}$ until the final value of 0.05 $\mathrm{cm^3}$ for Δ_Q was achieved in the fourth iteration. Once all subsets are localized, then the parameters are more constrained in the sampling procedure. This leads to more accepted runs in the next iteration.

Normally, sensitivity analyses are used to calculate sensitivity coefficients that characterize the behavior of the objective function at a particular location in parameter space, presumably in the vicinity of the true parameter values (Kool and Parker, 1988; Šimůnek and van Genuchten, 1996). In contrast to sensitivity analysis, PIMLI uses the variability in time of this sensitivity, expressed as the information content of a measurement with respect to the various model parameters, to select subsets of data that each contain explicit information for a particular parameter. Once identified, these subsets are used to constrain its corresponding parameter.

RESULTS AND DISCUSSION

Response Surfaces

As an example, contours of the objective function (Eq. [5]) for a sandy soil in six two-dimensional parameter planes are presented in Fig. 2. The six selected response surfaces are a representative set of the total 15 response surfaces. Most of the 15 response surfaces show relatively well-defined global minima and no additional local minima. However, the θ_r–n and K_s–λ response surfaces indicate a correlation between both parameters. If θ_r or K_s is perturbed positively, then n or λ will also be perturbed positively (Toorman et al., 1992). The relatively low sensitivity of the objective function for θ_r is due to the fact that the effective saturation (S_e) decreased only to about 0.08 within the range of measurements for the sandy soil used here. This makes the estimation of θ_r dependent on extrapolation beyond this range. A similar problem was also encountered by Parker et al. (1985) and Šimůnek et al. (1998). The relatively large area by the objective function = 0.10 contour in Fig. 2e and 2f indicates that there are many parameter combinations in the K_s–λ and θ_r–K_s parameter plane that all provide reasonably good predictions for the cumulative outflow, flux density, and the average water content curve in Fig. 1A. If a conventional parameter optimization technique is used to simultaneously estimate θ_s, θ_r, α, n, K_s, and λ based on the defined objective function, it cannot be successful because some of the response surfaces, like the K_s–λ and θ_r–K_s parameter planes, do not display a well-defined global minimum. Hence, it would depend on the sensitivity of the optimizer and the parameter start values, and it is unlikely to find a unique parameter set. The experimental work by Eching and Hopmans (1993) and Eching et al. (1994) showed that these uniqueness problems could be minimized if both cumulative outflow and soil water pressure head data were included in the objective function.

PIMLI

Figure 3 shows the information content of the measured cumulative outflow (Column A), water content (Column B), and flux density (Column C) of the sandy soil with respect to the six unsaturated flow parameters (Rows 1–6). PIMLI clearly shows that the saturated water content (θ_s) of the soil sample was the most sensitive parameter and that there is a distinct difference in information content of the various sets of measurements for the same parameter. Moreover, at the beginning of the outflow experiment, the average water content of the soil sample contains the most information for the θ_s parameter (Fig. 3B1). This is understandable because the water content is close to saturation at the beginning of the experiment and the sample is almost dry at the end. The distinct difference in information content of the average water content and cumulative outflow for the saturated water content is due to the fact that the cumulative outflow only provides information about a difference in water content and not about the absolute water content of the soil sample (i.e., cumulative outflow

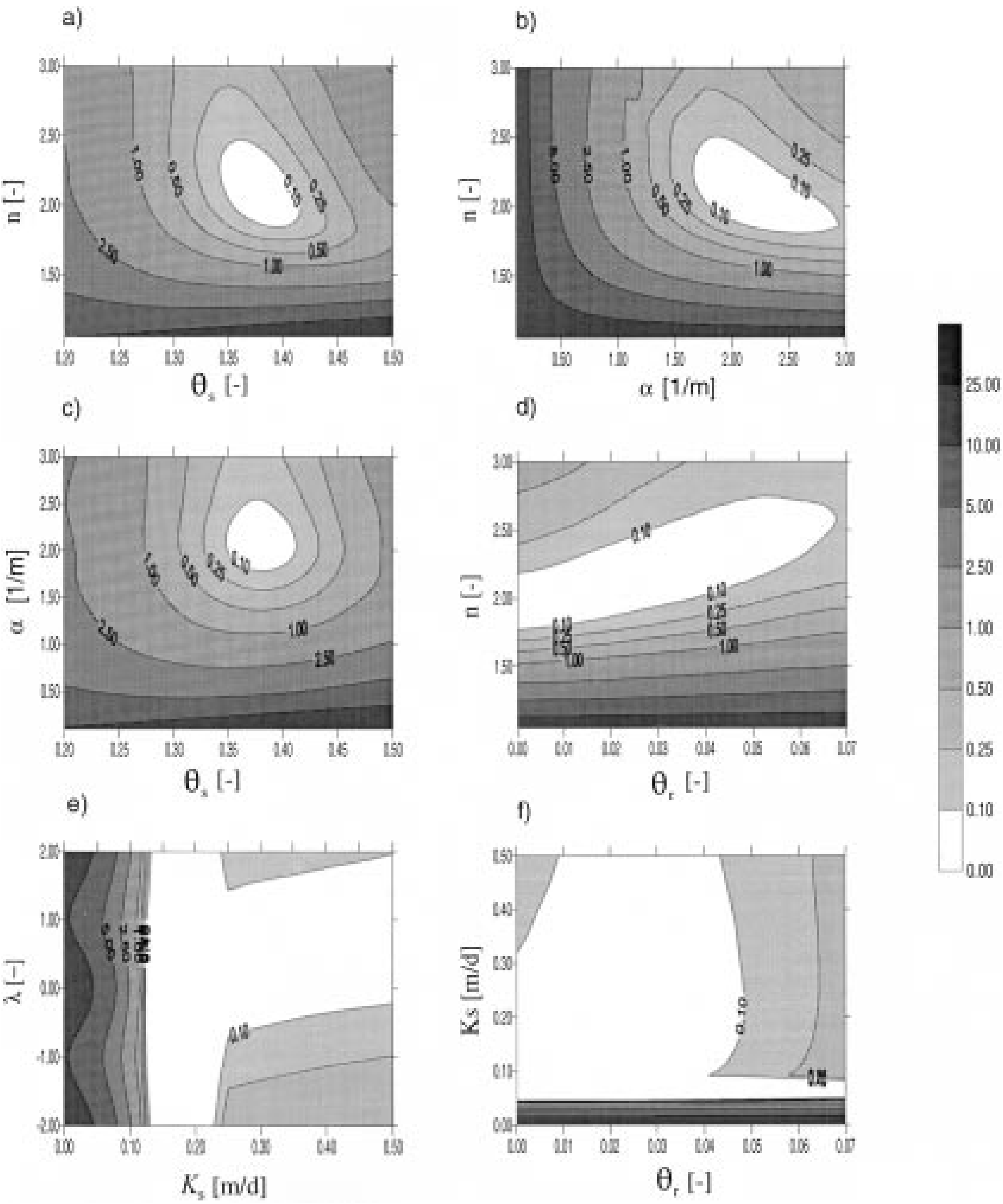

Fig. 2. Contours of the objective function OF(b) in the (a) θ_s–n, (b) α–n, (c) θ_s–α, (d) θ_r–n, (e) K_s–λ, and (f) θ_r–K_s parameter planes.

between $\theta_{begin} = 0.55$ and $\theta_{end} = 0.25$ is comparable with cumulative outflow between $\theta_{begin} = 0.35$ and $\theta_{end} = 0.05$). Simultaneous estimation of θ_s and θ_r on the basis of cumulative outflow only is therefore impossible (Van Dam et al., 1992).

Once the subset of θ_s is identified ($0 \leq t \leq 0.5$) and the histogram of the accepted parameters in this set is used for sampling in the next iteration, the information for the parameter α at hydraulic equilibrium in the cumulative outflow measurements at the beginning of the experiment becomes more apparent (Fig. 3A2). Thus, a pressure step that passes the air-entry value of the soil sample contains most information for the parameter α. The cumulative outflow measurements during hydraulic equilibrium after the first pressure increment ($1.0 \leq t \leq 1.5$) are therefore used to constrain α in the next iteration. When the uncertainty in the hydraulic parameters θ_s and α diminishes, then the information for the parameter n appears at hydraulic equilibrium in the cumulative outflow at intermediate pressure steps in the next iteration ($5.0 \leq t \leq 5.5$). Although the error in water content is much larger than the error in outflow, the water content of the soil sample still provides reasonable information for the parameter n.

In Iteration 3, after constraining θ_s, α, and n in the sampling procedure, the information content for λ is

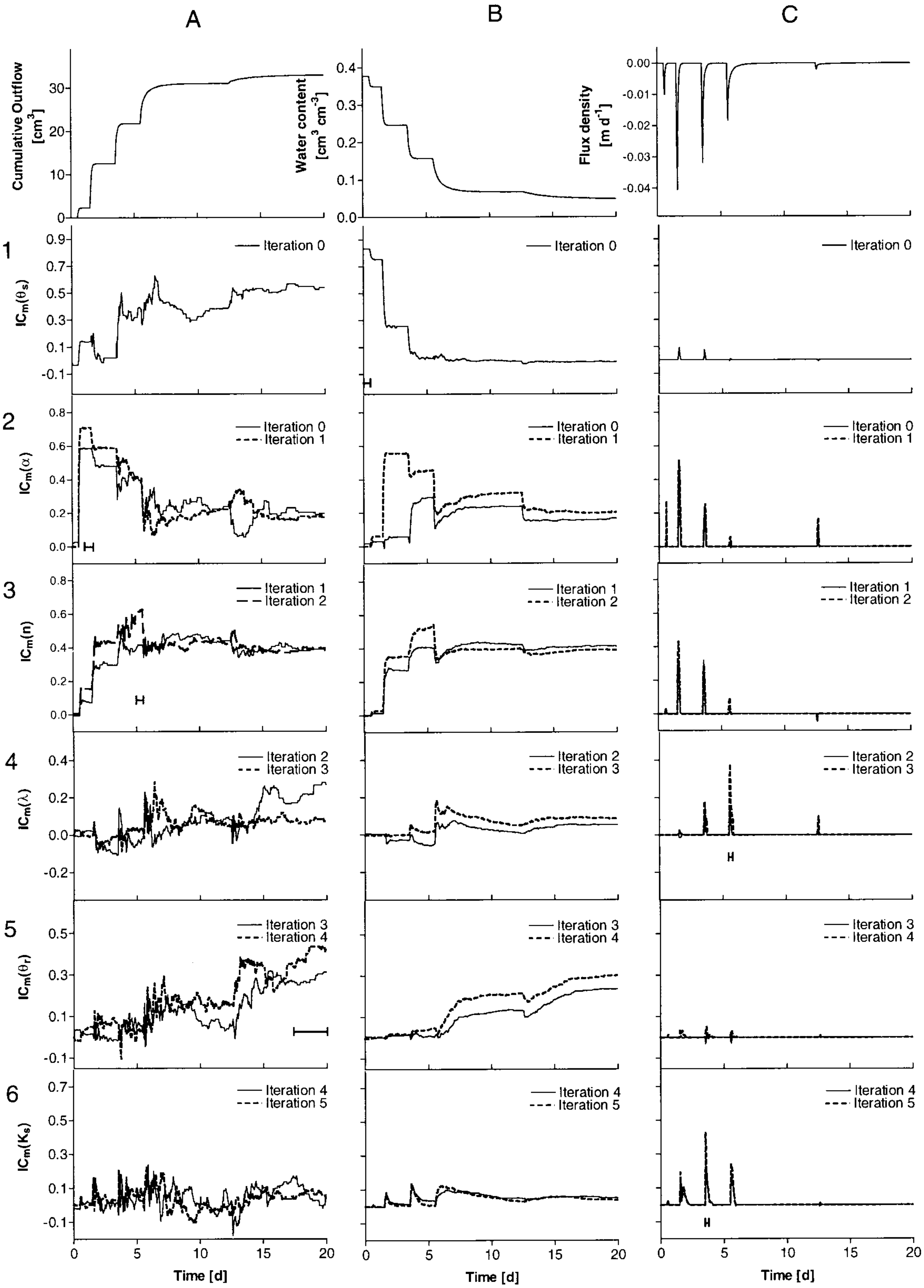

Fig. 3. Information content of (A) cumulative outflow, (B) water content, and (C) flux density measurements for the six Mualem–van Genuchten parameters (1–6) at different iterations. For a full explanation see text.

highest immediately after pressure increments in the low water content range of the soil sample. This makes sense because the factor S_e^λ in the hydraulic conductivity function, Eq. [3], is most dominant at lower water contents. Hence, the flux density measurements between $t = 5.5$ and $t = 5.75$ are used to constrain the parameter λ in the next iteration. Once the λ parameter is adjusted in the next iteration then the information of the residual water content (θ_r) appears in the final cumulative outflow at the end of the experiment ($18 \leq t \leq 20$). Hence,

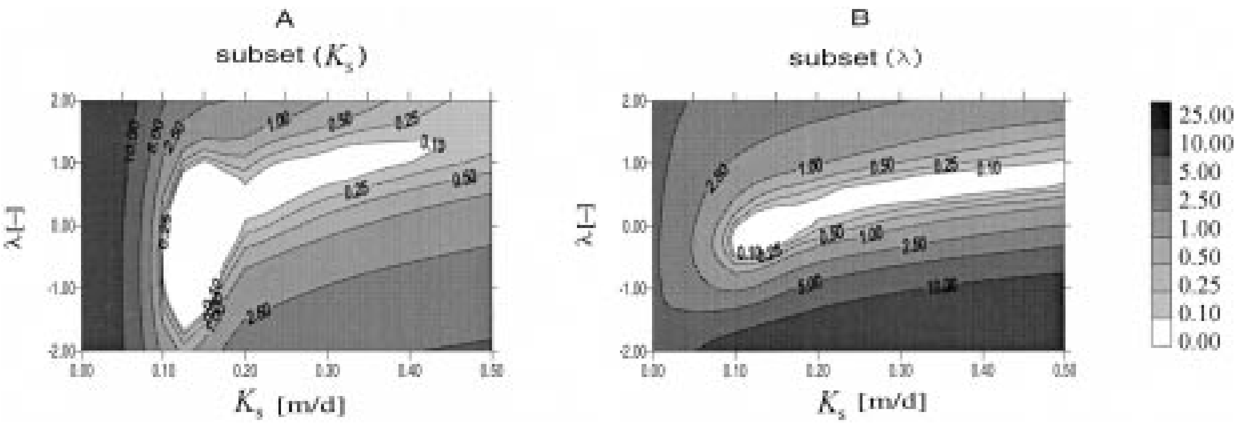

Fig. 4. Contours of the objective function in the K_s–λ parameter plane calculated for the selected subsets of measurements as distinguished with PIMLI.

if the shoulder and transition part of the water retention (Hollenbeck and Jensen, 1998) are constrained, then the final cumulative outflow of the soil sample provides a reasonable estimate for the parameter θ_r. If the error in water content would be reasonably lower than 0.021 $cm^3\ cm^{-3}$, the average water content of the soil sample can be used to further constrain θ_r. Once the uncertainty in the water retention of the soil sample and λ are reduced, the information for the saturated conductivity (K_s) appears in the flux density out of the soil sample (Fig. 3C6) after the third suction step ($2.0 \leq t \leq 2.25$). In general, the information for the parameters λ and K_s is not found in the entire time series of cumulative outflow or water content, but only in the flux density measurements immediately after a suction increment. This is also demonstrated if we compare the response surfaces of Fig. 2e with the response surfaces that are calculated with the selected sets of measurements for λ and K_s in the objective function (Fig. 4). The mutual dependency of the parameters K_s and λ, as found in case of the entire time series, is reduced if we focus on disjunctive subsets. The sensitivity of the OF(b) to λ strongly increases and the identifiability of λ therefore increases (Fig. 4B). As a result of this the identifiability of K_s also increases. Although not presented, the θ_r–n and θ_r–K_s parameter planes yield identical results.

Figure 5 presents the information content of the selected subsets of the various parameters as function of iteration number. It is clear that after each iteration, the information content for each parameter increases. This means that the parameter values converge. Moreover, for the parameters θ_s, α, and n, 10 iterations of 1000 Monte Carlo simulations are sufficient to accurately identify these flow parameters, whereas for K_s, λ, and θ_r, more iterations still improve the results.

Table 2 presents the original parameters of the reference run (sandy soil) and the mean and standard deviation of the hydraulic parameters after 20 iterations. We also included the results of PIMLI for a silty and clayey soil. From this table, it appears that PIMLI uniquely identifies the soil hydraulic parameters θ_s, α, n, K_s, and λ for a sandy and silty soil. The parameter values of the sandy and silty soil as used in the reference run lie within the 95% confidence interval as found with PIMLI. Moreover, standard deviations of parameter ranges are reduced with at least 75% for θ_r (silt) to 99% for n (silt). Identification problems only occur for λ (silt) and various parameters of the clayey soil where the effective saturation only decreased to 0.70 and therefore the range of measurements is not sufficient to select truly disjunctive subsets of measurements. Parameters then remain too much correlated, as found with the response surfaces in Fig. 2d. The overestimation of θ_r is then compensated by a slight overestimation of n. The relatively large standard deviation of λ for the clayey soil and partly the silty soil is due to the fact that most information for this parameter occurs at the dry end (see Fig. 3C4), also beyond the range of measurements of both soils.

Finally, in Fig. 6 we show the water retention curve of the clayey soil used in the reference run and the water retention as found with PIMLI after 20 iterations. It is clear that both water retention curves are identical within the range of measurements of this MSO experiment [$0.4 \leq \log(-h) \leq 2.7$]. Divergence occurs only beyond the range of measurements. This inverse problem is therefore ill posed or nonidentifiable because more than a single parameter set leads to the same model response within the range of measurements of this MSO experiment. This points at the major limitation for parameter identification. Only unique parameter estimates can be obtained if the range of measure-

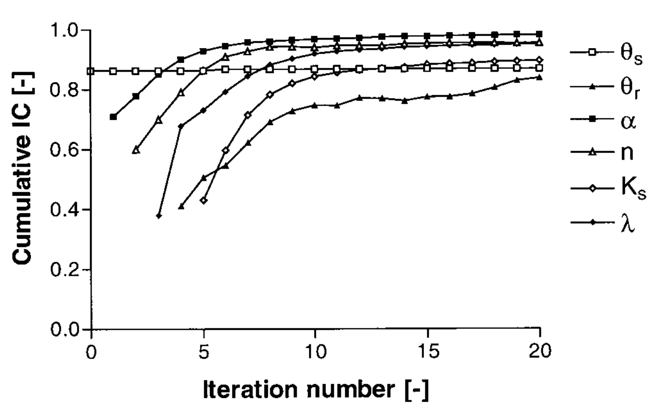

Fig. 5. Information content as function of iteration number for the different soil hydraulic parameters of the sandy soil.

Table 2. Soil hydraulic parameters for the sandy, silty, and clayey soil and the mean and standard deviation (in parentheses) for the disjunctive subsets as found with PIMLI after 20 iterations (italic).

Soil	θ_s	θ_r	α	n	K_s	λ
Sand	0.38	0.020	2.140	2.075	0.1556	0.039
PIMLI	*0.38 (0.011)*	*0.021 (0.003)*	*2.148 (0.016)*	*2.092 (0.025)*	*0.1714 (0.015)*	*0.104 (0.058)*
Silt†	0.46	0.034	1.600	1.370	0.0600	0.500
PIMLI	*0.46 (0.011)*	*0.029 (0.005)*	*1.600 (0.019)*	*1.361 (0.006)*	*0.0733 (0.008)*	*1.039 (0.560)*
Clay‡	0.42	0.000	1.910	1.152	0.1380	−1.384
PIMLI	*0.42 (0.011)*	*0.046 (0.006)*	*1.901 (0.041)*	*1.177 (0.003)*	*0.1415 (0.022)*	*0.300 (1.087)*

† Adapted from Carsel and Parish (1988).
‡ Adapted from Wösten et al. (1994).

ments is sufficient to divide the entire data set into disjunctive subsets that each contain the most information for the different unknown parameters. So, with this MSO experiment, we showed that there is enough information in the cumulative outflow, flux density, and water content to enable a unique parameter combination with PIMLI, if the range of measurements is large enough.

As PIMLI only uses a relatively small number of measurements with highest information content for a specific parameter, problems with the applicability of the presented method might occur when using PIMLI on real laboratory or field data sets that are corrupted with errors. Nevertheless, these problems can be solved by using an average of a number of measurements as a subset instead of all the single measurements.

CONCLUSIONS

Response surface analysis of a sandy soil showed that a conventional parameter optimization technique would experience problems with the simultaneous identification of θ_s, θ_r, α, n, K_s, and λ on the basis of cumulative outflow, water content, and flux density measured during a multi-step outflow experiment. Therefore, we propose PIMLI, a Parameter Identification Method based on the Localization of Information, that uses the variation in time of the model sensitivity for the various parameters to split the total set of measurements into disjunctive subsets that each contain the most information on one of the model parameters.

Results of PIMLI analysis showed that the average water content of the soil sample contains the most information for the saturated water content at the beginning of the outflow experiment. The distinct difference in information content of the average water content and cumulative outflow for the saturated water content is due to the fact that the cumulative outflow only provides information about a difference in water content and not about absolute water content in the soil sample. Simultaneous estimation of the saturated and residual water content on the basis of cumulative outflow only is therefore impossible (Van Dam et al., 1992). The cumulative outflow at hydraulic equilibrium after the pressure step that passes the air-entry value of the soil sample contains the most information for the parameter α. Later in the experiment, also during hydraulic equilibrium, the information for n is highest. Once the water retention of the soil sample is more constrained, then the information for the parameters λ and K_s appears in the flux density of the outflow of the soil sample. The cumulative outflow measurements at the end of the experiment contain the most information for the residual water content. The information content for K_s and λ is highest immediately after a pressure step in the high and low water content range, respectively. PIMLI analysis also shows the limitations of experimental data of which the range of measurements is not large enough to split the data set into truly disjunctive subsets.

Using the different localized subsets in an iterative Monte Carlo simulation procedure, we showed, using PIMLI, that there is enough information in the cumulative outflow, water content and flux density to enable a unique set of soil hydraulic parameters (θ_s, θ_r, α, n, K_s, and λ), if the range of measurements is sufficient.

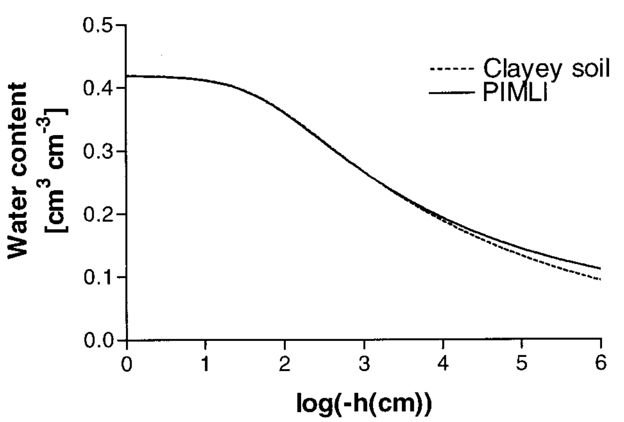

Fig. 6. Water retention curve of the clayey soil used in the reference run and the retention curve as found with PIMLI after 20 iterations.

ACKNOWLEDGMENTS

The current version of this manuscript has been greatly improved as a result of constructive review and comments by A. Sevink, J.W. Hopmans, E. Rastetter, J.M. Verstraten, and three anonymous reviewers. The investigations were (in part) supported by The Earth Life Sciences and Research Council (ALW), with financial aid from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Bohne, K., C. Roth, F.J. Leij, and M.Th. van Genuchten. 1993. Rapid method for estimating the unsaturated hydraulic conductivity from infiltration measurements. Soil Sci. Soc. Am. J. 155:237–244.

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

Celia, M.A., E.T. Bouloutas, and R.L. Zarba. 1990. A general mass-conservative numerical solution for the unsaturated flow equation. Water Resour. Res. 26:1483–1496.

Clausnitzer, V., and J.W. Hopmans. 1995. Non-linear parameter estimation: LM_OPT. General-purpose optimization code based on the Levenberg-Marquardt algorithm. Land, Air and Water Resources Paper 100032. Univ. of California, Davis.

Duan, Q., S. Sorooshian, and V.K. Gupta. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28:1015–1031.

Eching, S.O., and J.W. Hopmans. 1993. Optimization of hydraulic functions from transient outflow and soil water pressure head data. Soil Sci. Soc. Am. J. 57:1167–1175.

Eching, S.O., J.W. Hopmans, and O. Wendroth. 1994. Unsaturated hydraulic conductivity from transient multi-step outflow and soil water pressure data. Soil Sci. Soc. Am. J. 58:687–695.

Gupta, V.K., and S. Sorooshian. 1985. Uniqueness and observability of conceptual rainfall-runoff model parameters. The percolation process examined. Water Resour. Res. 19:269–276.

Gupta, H.V., S. Sorooshian, and P.O. Yapo. 1998. Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour. Res. 34:751–763.

Hollenbeck, K.J., and K.H. Jensen. 1998. Experimental evidence of randomness and non-uniqueness in unsaturated outflow experiments designed for hydraulic parameter estimation. Water Resour. Res. 34:595–602.

Hopmans, J.W., and J. Šimůnek. 1999. Review of inverse estimation of soil hydraulic properties. p. 643–659. *In* M.Th. van Genuchten et al. (ed.) Characterization and measurement of the hydraulic properties of unsaturated porous media. Univ. of California, Riverside.

Hudson, D.B., M.H. Young, P.J. Wierenga, and R.G. Hills. 1991. Transient method for estimating soil water parameters of unsaturated soil cores. p. 221. *In* Agronomy abstracts. ASA, CSSA, and SSSA, Madison, WI.

Janssen, P.H.M., and P.S.C. Heuberger. 1995. Calibration of process-oriented models. Ecol. Modell. 83:55–66.

Kool, J.B., J.C. Parker, and M.Th. van Genuchten. 1985a. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation. I. Theory and numerical studies. Soil Sci. Soc. Am. J. 49:1348–1354.

Kool, J.B., J.C. Parker, and M.Th. van Genuchten. 1985b. Onestep: A nonlinear parameter estimation program for evaluating soil hydraulic properties from one-step outflow experiments. Bull. 85-3. Virginia Agric. Exp. Stn., Blacksburg.

Kool, J.B., and J.C. Parker. 1988. Analysis of the inverse problem for transient unsaturated flow. Water Resour. Res. 24:817–830.

Kuczera, G. 1982. On the relationship of the reliability of parameter estimates and hydrological time series data used in calibration. Water Resour. Res. 18:146–154.

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

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

Musters, P.A.D., and W. Bouten. 2000. A method for identifying optimum strategies of measuring soil water contents for calibrating a root water uptake model. J. Hydrol. 227:273–286.

Parker, J.C., J.B. Kool, and M.Th. van Genuchten. 1985. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: II. Experimental studies. Soil Sci. Soc. Am. J. 49:1354–1359.

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

Šimůnek, J., and M.Th. van Genuchten. 1996. Estimating unsaturated soil hydraulic properties from tension disc infiltrometer data by numerical inversion. Water Resour. Res. 32:2683–2696.

Šimůnek, J., O. Wendroth, and M.Th. van Genuchten. 1998. Parameter estimation analysis of the evaporation method for determining soil hydraulic properties. Soil Sci. Soc. Am. J. 62:894–905.

Sorooshian, S., V.K. Gupta, and J.L. Fulton. 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall runoff models: Influence of calibration data variability and length on model credibility. Water Resour. Res. 19:251–259.

Tiktak, A., and W. Bouten. 1992. Modeling soil water dynamics in a forested ecosystem. III: Model description and evaluation of discretization. Hydrol. Proc. 6:455–465.

Toorman, A.F., P.J. Wierenga, and R.G. Hills. 1992. Parameter estimation of hydraulic properties from one-step outflow data. Water Resour. Res. 28:3021–3028.

Van Dam, J.C., J.N.M. Stricker, and P. Droogers. 1992. Inverse method for determining soil hydraulic functions from one-step outflow experiments. Soil Sci. Soc. Am. J. 56:1042–1050.

Van Dam, J.C., J.N.M. Stricker, and P. Droogers. 1994. Inverse method to determine soil hydraulic functions from multi step outflow experiments. Soil Sci. Soc. Am. J. 58:647–652.

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

van Genuchten, M.Th., F.J. Leij, and S.R. Yates. 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91/065. USEPA, Ada, OK.

Wösten, J.H.M., G.J. Veerman, and J. Stolte. 1994. Water retention and hydraulic conductivity characteristics of top- and subsoils in the Netherlands. The Staringreeks, Technical Document 18:1-66. DLO-Center, Wageningen, the Netherlands.

Yapo, P., H.V. Gupta, and S. Sorooshian. 1996. Calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. J. Hydrol. 181:23–48.

Zachmann, D.W., P.C. DuChateau, and A. Klute. 1981. The calibration of the Richards flow equation for a draining column by parameter identification. Soil Sci. Soc. Am. J. 45:1012–1016.