

Estimating effective model parameters for heterogeneous unsaturated flow using error models for bias correction

D. Erdal,¹ I. Neuweiler,¹ and J. A. Huisman²

Received 20 June 2011; revised 11 May 2012; accepted 14 May 2012; published 26 June 2012.

[1] Estimates of effective parameters for unsaturated flow models are typically based on observations taken on length scales smaller than the modeling scale. This complicates parameter estimation for heterogeneous soil structures. In this paper we attempt to account for soil structure not present in the flow model by using so-called external error models, which correct for bias in the likelihood function of a parameter estimation algorithm. The performance of external error models are investigated using data from three virtual reality experiments and one real world experiment. All experiments are multistep outflow and inflow experiments in columns packed with two sand types with different structures. First, effective parameters for equivalent homogeneous models for the different columns were estimated using soil moisture measurements taken at a few locations. This resulted in parameters that had a low predictive power for the averaged states of the soil moisture if the measurements did not adequately capture a representative elementary volume of the heterogeneous soil column. Second, parameter estimation was performed using error models that attempted to correct for bias introduced by soil structure not taken into account in the first estimation. Three different error models that required different amounts of prior knowledge about the heterogeneous structure were considered. The results showed that the introduction of an error model can help to obtain effective parameters with more predictive power with respect to the average soil water content in the system. This was especially true when the dynamic behavior of the flow process was analyzed.

Citation: Erdal, D., I. Neuweiler, and J. A. Huisman (2012), Estimating effective model parameters for heterogeneous unsaturated flow using error models for bias correction, *Water Resour. Res.*, 48, W06530, doi:10.1029/2011WR011062.

1. Introduction

[2] Modeling of water fluxes in the unsaturated zone is important for quantifying soil moisture movement between the surface and groundwater. This modeling is intrinsically difficult because the processes are highly nonlinear and soil structures can vary from millimeters to kilometers in size and can rarely be fully resolved. Animals and plants have a large impact on the topsoil, processes such as hysteresis and macropore transport may not be included in the model, measurement devices have errors and typically there are discrepancies between observation scale and modeling scale. Despite all these potential sources of errors, today we have advanced models that are assumed to adequately represent water flow in the unsaturated zone.

[3] A crucial point in modeling is to decide to what level the details of the system need to be resolved. A high level of detail may provide a better representation of reality, but it requires more data, system knowledge, and computational

power. Simpler models, on the other hand, require less details, are faster to run, and easier to understand, but may not accurately reproduce the system of interest. A decision on the appropriate level of detail depends on the modeling goal and the available data. Ideally, data and model simulations should be on the same scale, and the model should be able to represent the relevant processes. In modeling of the unsaturated zone for large scale systems, this is, however, rarely the case and the scale differences between observations and models can be orders of magnitude, as demonstrated for example by Vereecken *et al.* [2007]. To resolve differences in scale and to accurately describe spatially distributed processes of soil moisture flow and states, models need to be upscaled.

[4] In their review of upscaling methods, Vereecken *et al.* [2007] distinguished two ways of upscaling. The first way uses small scale spatial information to derive effective equations and/or effective parameters for the large scale model. Examples of such approaches are the use of stochastic theory [see, for example, Vereecken *et al.*, 2007; Zhang, 2002] and the scaleway approach of Vogel and Roth [2003]. The second way is to assume that the model equations can represent the effective behavior of the system and to estimate effective parameters using inverse modeling. Standard models for water flow in the unsaturated zone are often based on the Richards equation, also for larger scales. Although inverse methods to estimate effective parameters are well accepted, the assumption that such models can

¹Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, Hannover, Germany.

²IBG-3: Agrosphere, Forschungszentrum Jülich, Jülich, Germany.

Corresponding author: D. Erdal, Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, D-30167 Hannover, Germany. (erdal@hydromech.uni-hannover.de)

represent the effective behavior is not always physically justified [e.g., Vereecken *et al.*, 2007]. For this and several other reasons, the estimation of effective parameters is notoriously difficult in the unsaturated zone, as evidenced by the studies of Papafiotiou *et al.* [2008], Mertens *et al.* [2005], and Kumar *et al.* [2010] that all showed differences between parameters estimated for the same system at different scales.

[5] A common approach to estimate hydraulic parameters of small soil columns is to use one- or multistep outflow experiments, in which a saturated soil sample is drained by (stepwise) lowering of the water pressure at the bottom of the sample [e.g., Bayer *et al.*, 2005; Laloy *et al.*, 2010a; Schelle *et al.*, 2010; Valiantzas and Londra, 2008; Vasin *et al.*, 2008; Zurmühl and Durner, 1998]. The ability to obtain representative hydraulic parameters for such an experiment depends, naturally, on the level of complexity of the system [Laloy *et al.*, 2010a; Vasin *et al.*, 2008], the type of model used [Laloy *et al.*, 2010a; Zurmühl and Durner, 1998], but also to a large extent on what system states the model seeks to reproduce. For example, Bayer *et al.* [2005], Durner *et al.* [2008], and Laloy *et al.* [2010a] all found poor agreement between measured and modeled internal water distribution, even though the averaged state as determined by the flow in and out of the system was well described. The difficulties of estimating effective hydraulic parameters have also been pointed out by Vereecken *et al.* [2008], who reviewed problems with estimating hydraulic properties from soil moisture data and by Schelle *et al.* [2010], who showed that it is particularly difficult to predict the soil moisture regime outside the calibration period used to estimate the effective parameters.

[6] All papers discussed in the previous paragraph ask the question if, or to what extent, it is possible to find effective model parameters that reproduce the observations well. If one believes that representative effective parameters exist, it remains a challenging task to find them. Typically, an optimization algorithm is used to search the parameter space and to find the best possible parameter combination that minimizes the difference between observations and model predictions. A major problem in inverse modeling applications that estimate hydraulic parameters of the unsaturated zone is the long run time of a single flow model evaluation, which restricts our ability to adequately explore the parameter space. This, in turn, can lead to an additional uncertainty in the resulting parameter estimates. With increasing computational power, methods for automatic model parameter estimation have become increasingly popular for estimating parameters within an acceptable number of model evaluations. Apart from estimating the best set of parameters for a particular problem, some methods also assess model parameter uncertainty. Examples of such methods are Markov chain Monte Carlo (MCMC) methods [e.g., Gelman *et al.*, 2004; Vrugt *et al.*, 2008], informal Bayesian approaches using generalized likelihood functions [e.g., Beven and Freer, 2001], and multiobjective parameter estimation approaches that search for an entire set of solutions that are all optimal in the sense that an improvement in one objective results in a deterioration of another objective [e.g., Vrugt *et al.*, 2003].

[7] An important decision in setting up an inverse modeling problem is the definition of a likelihood function because this automatically entails assumptions about the

underlying causes for the difference between observations and model. In the theory of using a model, Kennedy and O'Hagan [2001] discussed six groups of errors that might cause deviations between models and observations: parameter uncertainty, model inadequacy, residual variability, parametric variability, measurement error, and code uncertainty. Of these errors, the measurement error is the most commonly treated. It is often assumed that it can be treated as uncorrelated noise that follows a Gaussian distribution with zero expectation (i.e., white noise), which allows treatment with well-established statistical methods. The uncertainty resulting from a possibly incomplete search of the parameter space, and hence the risk of only finding a local optimum, is in this view a code uncertainty. In unsaturated zone modeling, it is common that certain structures of the soil or certain processes are not well represented, which makes the model an imperfect model of the real world. This is referred to as model inadequacy. Due to the often strong correlation in time and space, it is inappropriate to describe these two error sources as uncorrelated Gaussian noise. Therefore, it is common practice to ignore errors due to model inadequacy and code uncertainty, despite the fact that these errors can be orders of magnitude larger than measurement errors [Doherty and Welter, 2010].

[8] An alternative approach to deal with errors in modeling is to include external error models that correct for discrepancies between observations and modeling predictions. Examples of such approaches are the use of autoregressive models and external adjustments of model forcing terms used by Kavetski *et al.* [2006], Vrugt *et al.* [2008], and Reichert and Meileitner [2009]. Recently, a formal Bayesian approach was proposed that uses a likelihood function that can take into account skewness, heteroscedasticity, and correlation of the residuals [Schoups and Vrugt, 2010]. This method was successfully tested for a hydrological test case. Depending on the complexity of a problem and the availability of data, different problems might require different error treatments. In this context, Doherty and Welter [2010] pointed out that a universal procedure to deal with modeling errors associated with imperfect models does not exist.

[9] In this paper we aim to investigate how parameters can be estimated for a model that is known to be imperfect because it does not fully resolve soil structure. The motivation to address this question is that measurements are regularly made on a much smaller scale than the modeling scale. For example, effective hydraulic parameters for large scale models describing water flow in the unsaturated zone are commonly estimated from water content observations made with TDR probes on a centimeter scale. We follow the general idea of introducing an error model to the parameter estimation process such as described by Carter [2004]. We test the approach by using spatially averaged saturation measurements taken during multistep outflow experiments in lab-scale heterogeneous sand samples by Vasin *et al.* [2008]. The available data were obtained using neutron radiography, neutron tomography, and outflow measurements. We would like to stress here that these data are used for illustrative purposes, but that the presented ideas are by no means limited to this type of experiment. In fact, our results are not meant to improve multistep outflow experiments, but instead they are intended for use with a wide range of

hydrological flow problems, where typical applications would be very different from this small scale.

[10] The remainder of the paper is structured as follows. First, the setup of the experiment of *Vasin et al.* [2008] and a virtual reality experiment used for initial tests are explained in section 2. Section 3 presents a selection of parameter estimation results using MCMC simulation to illustrate the problem of estimating effective parameters for heterogeneous soils and the need for further investigation. In section 4, three external error models are introduced and tested as a way to improve model performance by accounting for soil structure outside of the flow model. Finally, conclusions are drawn in section 5.

2. Methods

2.1. Data and Measurements

[11] Two sources of data are used in this paper: data from a real multistep outflow (MSO) experiment (from here on referred to as RE) performed by *Vasin et al.* [2008], and data from simulated virtual reality drainage experiments (referred to as VR) based on the same concept as the real experiment. The experiment of *Vasin et al.* [2008] used a $10 \times 10 \times 20 \text{ cm}^3$ column for the MSO experiment. Neutron tomography was used to obtain the three-dimensional (3-D) water content distribution at hydrostatic equilibrium and radiography was used to get two-dimensional (2-D) (horizontally averaged) saturation distributions under dynamic flow conditions. Outflow measurements provided information about the total water mass balance. The column was packed heterogeneously with cubes of two different sand types with particle distributions of 0.08–0.2 and 0.1–0.5 mm. A periodic and a random structure of coarse sand inclusions in a fine sand matrix were created. Despite the advanced techniques used, the data still contain several sources of error (e.g., neutron scattering, uncertain transformation of measured beam intensity into water content, imperfect packing, artifacts due to imperfect contacts between the cubes). Therefore, simulated data are initially

used to develop the error model concepts in an environment that only considers model inadequacy as a source of error.

[12] The soil columns used in the MSO experiment were drained by a stepwise application of five pressure heads (–10, –20, –30, –40, –50 cm) at the bottom of the column. In between each pressure step, the column was allowed to reach hydrostatic equilibrium. The setup of each data set is further explained in section 2.2.1 and section 2.2.2. The heterogeneous packing of the soil columns consisted of 20 horizontal layers. Radiography measurements of the RE resulted in 2-D images of horizontally averaged water saturation (further information about the acquisition of the RE data, see *Vasin et al.* [2008]). To make the RE and the VR data similar, the 2-D RE data were horizontally averaged, which resulted in one-dimensional (1-D) profiles of 40 spatially (vertically) spread measurement points for each data set. Such a point is from here on referred to as a horizontally averaged layer (HAL) and the concept is further illustrated in Figure 1.

[13] As outlined above, the aim of our parameter estimation is to find an effective homogeneous model that represents the average saturation of the column (i.e., the change in storage due to infiltration and outflow) using a limited set of small-scale spatial measurements. To illustrate the problems and our ideas for solutions, five measurement strategies are considered in this paper. The five strategies use different combinations of local (horizontally averaged) measurements of average saturation taken at the same time and they are summarized in Table 1. The strategies can be divided into three groups. The first group of observation strategies uses measurements taken close to each other, hence seeing a smaller volume of the column but capturing larger structures. The second group of strategies uses measurements spread over the column, hence seeing more of the volume of the column but not capturing larger structures. The difference between these groups can be understood by considering the periodic structure in Figure 1 and comparing a strategy that sees the full top inclusion (small coverage, large structure) with one that sees only the top HAL,

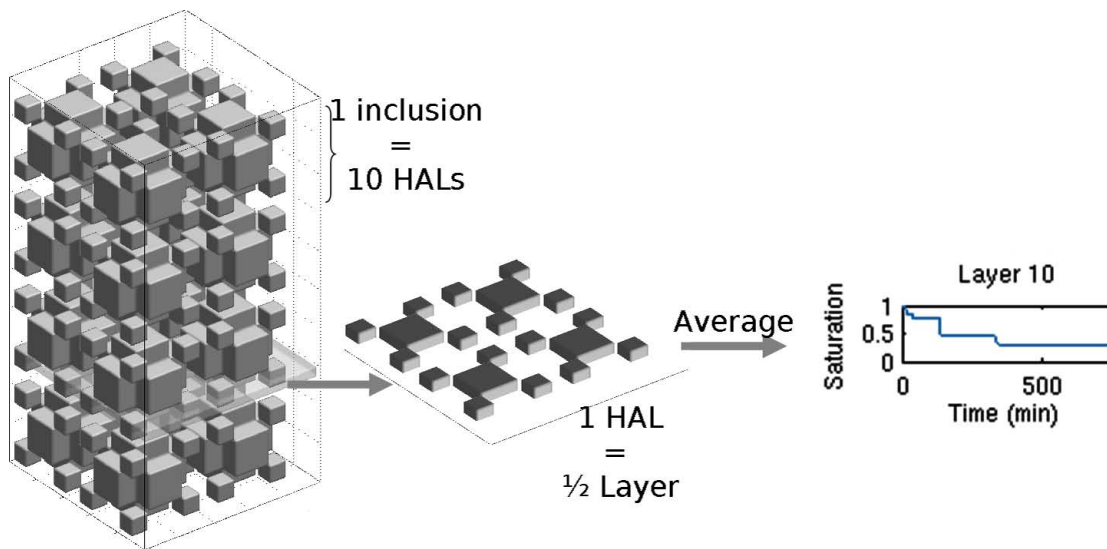


Figure 1. A horizontally averaged layer (HAL) and the size of an inclusion for the periodic structure. Gray parts indicate coarse sand while transparent parts indicate fine sand.

Table 1. The Different Measurement Strategies Used for the Parameter Estimations^a

Strategy	No. of HALs	HAL Used
All data	40	1, 2, 3, ..., 40
Top	5	35, 36, 37, 38, 39
Spread 1	5	5, 6, 17, 18, 30
Spread 2	5	5, 11, 15, 18, 30
Connected	5	20, 21, 22, 23, 24

^aThe horizontally averaged layers (HALs) are counted from bottom (1) to top (40) of the column (cf. Figure 1).

the bottom HAL, and one HAL in between (large coverage, no structure). The third group of strategies uses all available spatial and temporal data. Other approaches relying on heterogeneous models are available to analyze this third case, and they might provide better results. However, the focus of this paper is on strategies to improve effective parameter estimation and this reference case is presented for comparison only. The first group is represented by the top and the connected strategies, the second group is represented by two spread strategies and the last group is represented by the all data strategy (Table 1).

[14] This way of treating the data from a MSO experiment is unconventional and it is clear that spatially spread soil moisture measurements are typically not available in this type of experiment. However, the idea of the effective parameter estimation performed here should not be confused with the common use and goal of a MSO experiment. The aim of this paper is to investigate approaches to obtain an upscaled simple model based on local and spatially spread out measurements. We only use data from MSO experiments because of the availability of the RE data of *Vasin et al.* [2008], which allows very detailed investigations of our research questions. Clearly, the ideas explored here are meant to be applicable to a wider range of hydrological

flow problems, and surely extend beyond the MSO experiment discussed here.

2.1.1. Real Experiment Data

[15] The data presented by *Vasin et al.* [2008] consisted of two MSO experiments, one for a periodic structure and one for a random structure. These structures are shown in Figure 2 (left and middle) and consist of two sands arranged in different patterns. The two structures are considered because they have different representative elementary volumes (REV). The periodic structure has a perfect REV (one inclusion) but the volume percentage of coarse sand differs between the layers, while the random structure has no REV but instead has the similar volume percentage of coarse sand in each layer. We used horizontally averaged observations (2-D) from neutron radiography, which were taken once every minute during drainage after the water pressure at the bottom of the column was decreased. The drainage of the periodic structure resulted in strong air entrapment inside the coarse sand inclusions. Since the effects of air entrapment are not represented by the model used in our simulations, the periodic structure is only considered for the VR cases. More information about the experiment can be found by *Vasin et al.* [2008].

[16] To increase the number of real data cases, the random structure is analyzed in three different approaches. In the first approach we used the full data set with all horizontally averaged layers (1-D). The second and third approach use measurements from only two vertical stripes, called line A and line B (see Figure 2) that represent a much more layered structure than the fully averaged data of the entire random structure. Because of this more layered structure, line A and line B represent more difficult cases for effective parameter estimation.

[17] The MSO experiment carried out by *Vasin et al.* [2008] used five pressure heads, applied in sequence at the

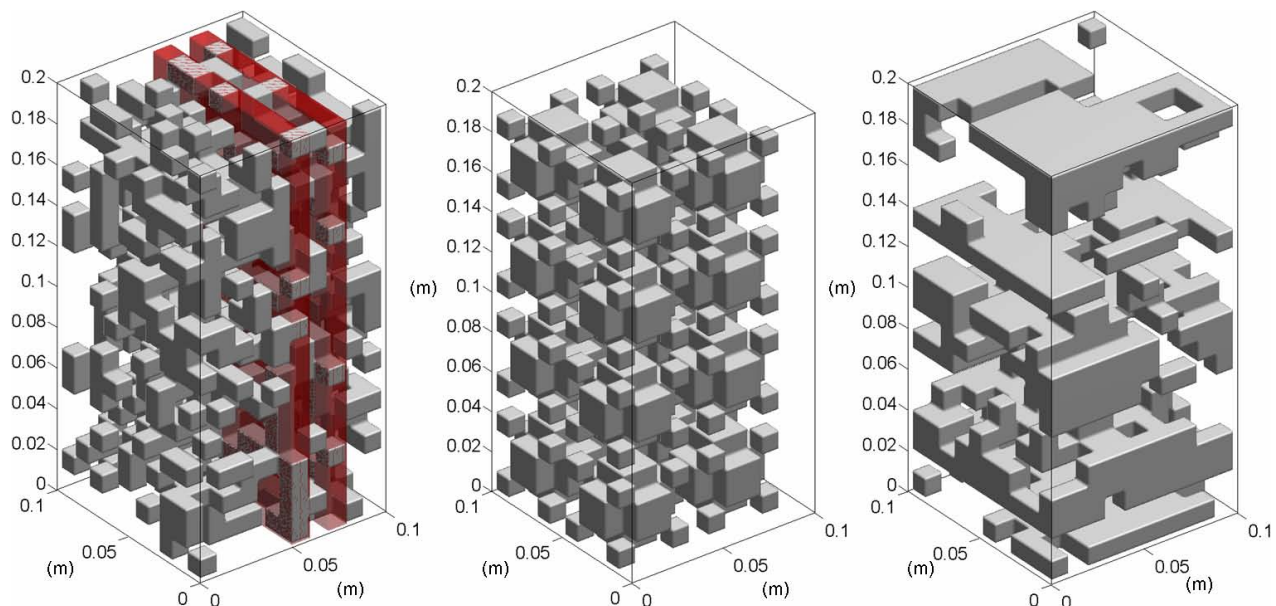


Figure 2. The structures used in the test cases: random (left), periodic (middle), and layered (right). Gray parts indicate the coarse sand while the fine sand is transparent. The red sections in the random structure highlight the subsections line A (left) and line B (right).

bottom of the column, while the side walls were closed and the top was open to the surrounding. Since the duration of the experiment was rather short (<24 h) evaporation from the open top was neglected. For the RE data, we use measurements at 113 points in time so that a full data set consists of 113 temporal \times 40 HAL measurement points. Of these 40 HALs, different combinations are used as observation data for effective parameter estimation, as explained in Table 1.

2.1.2. Virtual Reality Data

[18] For the generation of the heterogeneous virtual reality, three soil columns were used: a random and a periodic structure based on the RE and one extra column with a layered structure, which is also shown in Figure 2 (right). The layered structure is used specifically because of the unfavorable behavior caused by the variation in volume fraction of coarse and fine sand between each layer, which means that no REV can be found. Of course the layered structure is most representative for real soils and more likely to occur in the field than the periodic and the random structure.

[19] The hydrological model used in this study relies on the Richards equation:

$$\frac{n_f \partial S}{\partial t} - \nabla \cdot [K_u(h)(\nabla h + \mathbf{1})] = 0, \quad (1)$$

where $S(-)$ is the water saturation, $n_f(-)$ is the porosity, K_u (m s^{-1}) the unsaturated hydraulic conductivity, h (m) is the water pressure head (negative for unsaturated conditions), and $\mathbf{1}$ is the unit vector in z direction positive upward. For the water retention and hydraulic conductivity curves an approach similar to the Brooks-Corey parameterization [Brooks and Corey, 1966] is used, but with a different exponent for the hydraulic conductivity and retention curves:

$$S_e = \begin{cases} \frac{S - S_{rw}}{S_{sat} - S_{rw}} = \left(\frac{h_d}{h}\right)^{\lambda_{sat}} & \text{for } h > h_d \\ 1 & \text{for } h \leq h_d, \end{cases} \quad (2)$$

$$K_u = K_{sat} S_e^{\lambda_{perm}},$$

where $S_e(-)$ is the effective water saturation, $S_{rw}(-)$ is the residual water saturation, $S_{sat}(-)$ is the maximum water saturation, h_d (m) is the air entry pressure head, K_{sat} (m s^{-1}) is the saturated hydraulic conductivity, and $\lambda_{perm}(-)$ and $\lambda_{sat}(-)$ are shape parameters.

[20] All columns have a size of $10 \times 10 \times 20 \text{ cm}^3$ and consist of 2000 cells of size $1 \times 1 \times 1 \text{ cm}^3$ that each can be fine or coarse sand. For the simulations, the system is discretized into 16,000 cells ($20 \times 20 \times 40$). Using finer resolutions did not change the results significantly. Table 2 shows the hydraulic parameters used for the two sands.

Materials with a strong contrast were chosen to create a difficult case study and to clearly be able to distinguish between the two materials.

[21] The model ParFlow [Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet and Maxwell, 2005] is used for all water flow simulations in this study. ParFlow solves the Richards equation using an implicit backward Euler finite difference scheme and a Newton-Krylov non-linear solver. Simulations of the VR MSO experiment are set up to be similar to the real experiment. The same five pressure heads as in the real experiment are applied at the bottom of the column, while the side walls and the top are simulated as no flux boundaries. For the VR data, 68 data points are used in time, so that each column has a full data set of 68 temporal \times 40 HAL measurement points.

2.2. Parameter Estimation Setup

[22] The observation data from the virtual reality and the real experiments were used to estimate effective model parameters for a homogeneous model. Examples in the literature [e.g., Coppola et al., 2008; Vogel et al., 2008; Zurmühl and Durner, 1998] and early trials (not shown here) have shown that a single set of effective parameters cannot capture the main features of the outflow curve of such a heterogeneous column. Therefore, a simplified two-material formulation is used for the pressure-saturation relation:

$$S_e = \frac{S - S_{rw}}{S_{sat} - S_{rw}} = \Phi^{(coarse)} * S_e^{(coarse)} + \Phi^{(fine)} * S_e^{(fine)}, \quad (3)$$

$$K_u = K_{sat} S_e^{(\text{eff})} S_e^{\lambda_{perm}^{(\text{eff})}},$$

where $S_e(-)$ is the averaged effective saturation, $S_{sat}^{(A)}(-)$ and $S_{rw}^{(A)}(-)$ are the average maximum saturation and residual saturation, respectively, computed as $S_{sat/rw}^{(A)} = \Phi^{(coarse)} * S_{sat/rw}^{(coarse)} + \Phi^{(fine)} * S_{sat/rw}^{(fine)}$, $\Phi^{(A)}(-)$ is the volume ratio of material A, $S_e^{(A)}$ is the S_e value from equation (2) for material A, K_u is the average unsaturated hydraulic conductivity, and $\lambda_{perm}^{(\text{eff})}$ is a shape parameter for the averaged hydraulic conductivity curve. This formulation hence takes the weighted arithmetic mean of the retention curves of two materials that both have a set of Brooks-Corey parameters and is similar to the formulation used by Durner [1994]. When using equation (3) it is assumed that the two materials are separable and have a well defined volume fraction. It is also assumed that the two materials are connected throughout the structure to avoid local entrapment. A summary of the parameters used for effective parameter estimation and their prior distribution are given in Table 3.

[23] Equation (3) can be used to derive an upscaled model when the structure is known in detail, as is the case in this study. This has been discussed by Vasin et al. [2008]

Table 2. Parameter Values for the Two Sands^a

Material	K_{sat} (m s^{-1})	λ_{sat} (-)	λ_{perm} (-)	h_d (m)	Φ (-)	S_{sat} (-)	S_{res} (-)	n_f (-)
Coarse sand	5.83×10^{-2}	2.9	3.2	0.25	0.32	0.97	0.10	0.43
Fine sand	3.78×10^{-4}	4.7	3.2	0.44	0.68	0.97	0.14	0.43

^aFor explanation of the parameters see equation (2). Φ is the volume ratio of each material and n_f is the porosity.

Table 3. Prior Parameter Ranges for the Hydraulic Parameters^a

	$K_{\text{sat}}^{(\text{eff})}$	$\lambda_{\text{sat}}^{(\text{coarse})}$	$\lambda_{\text{sat}}^{(\text{fine})}$	$\lambda_{\text{perm}}^{(\text{eff})}$	$h_d^{(\text{coarse})}$	$h_d^{(\text{fine})}$	Φ
Upper limit	-2	3.4	3.4	3.9	0.9	0.9	0.9
Lower limit	-16	-0.7	-0.7	0	0.1	0.1	0.1

^a K_{sat} and the three λ values are log transformed to increase sampling efficiency.

for the random and the periodic structure. For the random structure, which shows only small variations in the volume percentage of coarse and fine sand between the layers, an excellent match to the assumptions of equation (3) is expected. However, the effective hydraulic conductivity function cannot easily be cast into a common parameterization valid for arbitrary structures. Therefore, we chose to use the simplified parameterization given in equation (3).

[24] The homogeneous model with effective hydraulic properties is solved in 1-D using ParFlow with 40 layers and a maximum time step of 60 s. Model output is generated with 60 s intervals. The seven hydraulic parameters (Table 3) are estimated using Markov chain Monte Carlo (MCMC) simulation using the DREAM algorithm [Vrugt *et al.*, 2008]. The DREAM algorithm is chosen because it is an efficient MCMC parameter estimation algorithm that provides both the best fitting parameters and their uncertainty. In this work we initially assume that the residuals r are mutually independent and follow a Gaussian distribution with zero mean and known standard deviation $\sigma = 0.15$. Our choice of σ reflects the large variations in saturation found in the full 3-D column and is assumed to integrate over several error sources, including those associated with using a homogeneous effective model to represent a heterogeneous reality. The chosen value may therefore seem large, but does well reflect the standard deviation of the saturation data. For a simulated saturation S_{sim} , obtained with the effective hydraulic parameters θ , the posterior probability density p , given observations S_{obs} , and a noninformative prior, can be calculated as [e.g., Gelman *et al.*, 2004, p. 48]:

$$r(\theta, S_{\text{obs}}) = S_{\text{obs}} - S_{\text{sim}}(\theta), \quad (4)$$

$$p(\theta|S_{\text{obs}}) \propto \exp \left[-\frac{1}{2} \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} \frac{r_{j,i}^2(\theta, S_{\text{obs}})}{\sigma^2} \right], \quad (5)$$

where N_t and N_s are the number of observations in time and space, respectively. The convergence of the MCMC simulation is assessed using the \hat{R} convergence criterium [e.g., Gelman *et al.*, 2004, p. 297]. All MCMC simulation results presented here are based on at least 1000 draws from

the posterior distribution after converge was achieved ($\hat{R} < 1.2$).

2.3. Validation

[25] Three validation scenarios were defined for the VR cases in order to assess the predictive power of the estimated hydraulic parameters. All three scenarios are evaluated using the average saturation of the full column ($\bar{S}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} S_{j,i}$) and the internal water distribution is not considered. The three scenarios are: the multistep drainage experiment used for the parameter estimation, a three step infiltration scenario, and a one-step drainage scenario. The initial and boundary conditions used in these validation scenarios are shown in Table 4 and the flow behavior can be seen in Figures 3, 4, and 5. The two new validation scenarios were selected to provide more insights in the reproduction of flow dynamics of the estimated effective hydraulic parameters.

[26] At this point, it is important to note that we are not interested in the actual parameter values resulting from parameter estimation in this study. Instead, we focus only on the predictive power of the estimated parameters in the validation scenarios. Evaluation of the predictive power is done in two ways. First, the maximum likelihood parameter set from the MCMC simulation is used to evaluate model performance by visual comparison of measured and modeled average saturation for the three validation scenarios. Second, the root mean square error (RMSE) is calculated to quantify model performance:

$$\text{RMSE} = \sqrt{\frac{1}{N_t} \sum_{j=1}^{N_t} g_j^2}. \quad (6)$$

The RMSE is calculated both from the difference in water saturation ($g_j = \bar{S}_{\text{obs},j} - \bar{S}_{\text{sim},j}$) and the difference in the rate of change of the saturation ($g_j = \partial \bar{S}_{\text{obs},j} / \partial t_j - \partial \bar{S}_{\text{sim},j} / \partial t_j$), where the required time derivative is calculated numerically using backward finite difference.

[27] For the RE, independent validation scenarios are not available. Although infiltration experiments were performed on the random structure, hysteresis effects were so strong that a meaningful evaluation is not possible in the context of this study. Therefore, the RE is only evaluated using the fully averaged data from the multistep outflow experiment. For clarity in the following figures, only the maximum likelihood model predictions are shown. The posterior distribution of the parameters and model predictions are illustrated in separate tables and figures.

Table 4. Setup of the Three Validation Scenarios^a

Name	Initial Condition	Top (mm s ⁻¹)	Boundaries Bottom (m)	Sides	Times (min)
Drainage	FS	NF	-0.1, -0.2, -0.3, -0.4, -0.5	NF	18, 35, 133, 333, 600
Infiltration	HE	0, -0.02, 0, -0.04, 0	-0.4	NF	10, 20, 40, 50, 70
One step	FS	NF	-0.4	NF	50

^aTimes refer to the time when a boundary is changed, and the last time is the duration of the simulation (FS = fully saturated, HE = hydrostatic equilibrium, NF = no flow).

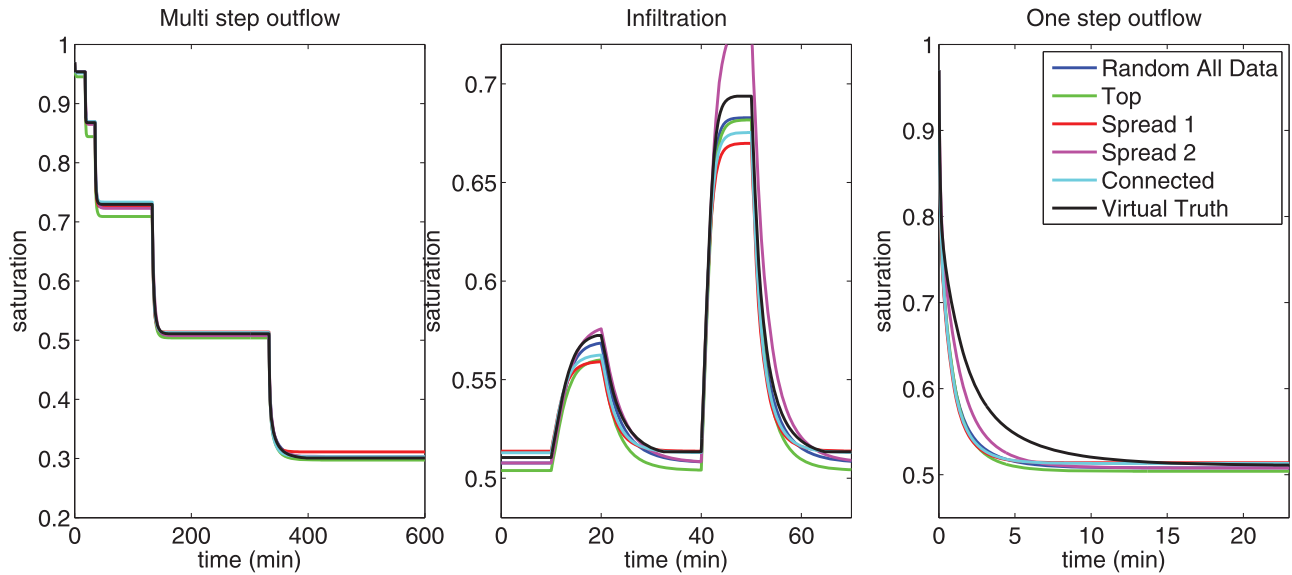


Figure 3. Results of the validation scenarios for the random structure using different measurement strategies.

3. Estimation of Effective Hydraulic Parameters

3.1. Virtual Reality

[28] Effective hydraulic parameters were estimated for a total of around 100 different simulation scenarios using different structures and different combinations of spatial data. The results show that the different structures performed very differently, as would be expected. For the random structure, the MCMC algorithm converged quickly and the resulting parameters performed well in the validation scenarios. The results for the periodic structure showed a strong dependence on where the measurements were taken in the column. Parameter estimation for the layered structure is more difficult compared to the other two structures

because of the complex structure. In the following, the performance of the estimated effective hydraulic parameters in the validation scenarios is described in detail for each structure. The validation scenarios are shown in Figures 3, 4, and 5 for the random structure, the periodic structure, and the layered structure, respectively. The associated model performance is summarized in Table 5.

[29] The estimated hydraulic parameters performed well for the random structure, which can be explained in two steps. First, the underlying assumptions of the two-material Brooks-Corey parameterization (equation (3)) are that each layer has the same volume fraction for the two materials and that each of the materials is connected throughout the

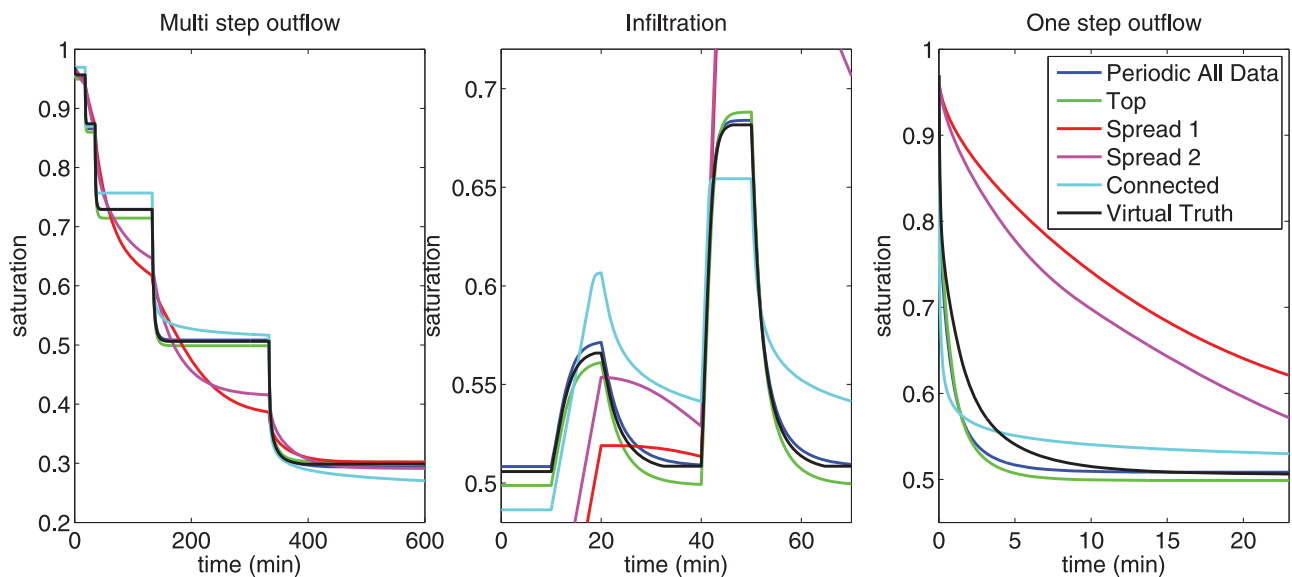


Figure 4. Results of the validation scenarios for the periodic structure using different measurement strategies.

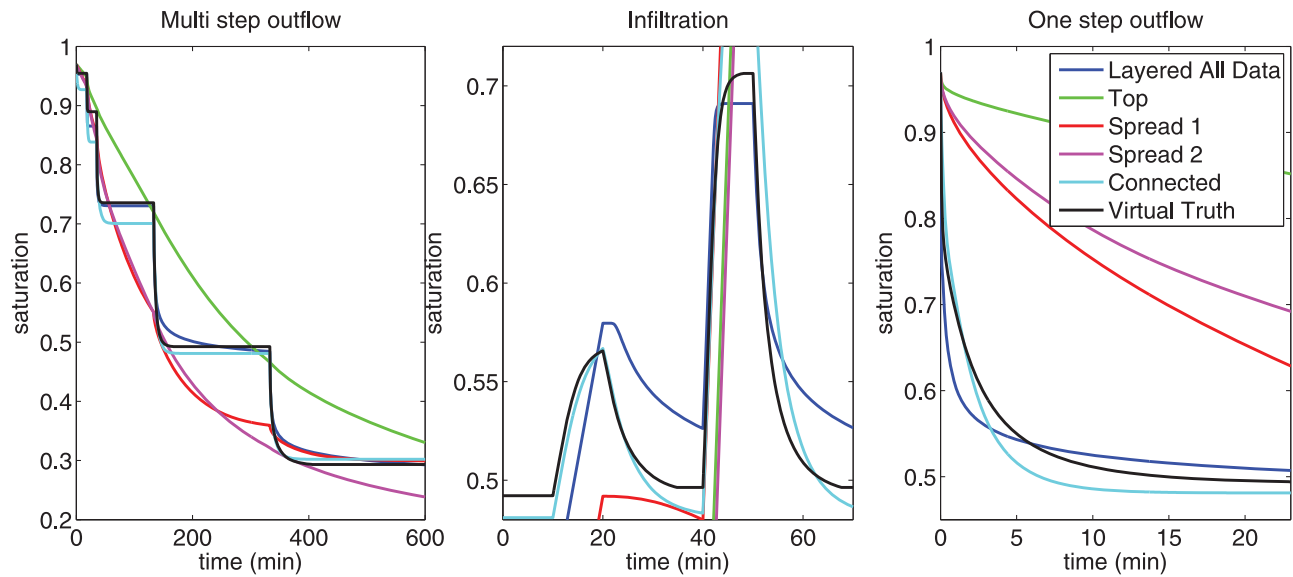


Figure 5. Results of the validation scenarios for the layered structure using different measurement strategies.

sample. These assumptions hold fairly well for the random structure. The second reason is that the REV used in the parameter estimation (one cell in 1-D) matches the REV of the horizontally averaged random structure. Because the ratio between the two materials in the random structure is close to 35% for most layers, each HAL is a good representation of the full system and it does not matter so much which layers are used or how they are combined.

[30] In the case of the periodic structure, the REV also plays an important role. In this structure, the REV is made up of the 10 layers that cover one inclusion (Figure 1). The validation results for the periodic structure show that measurement strategies that capture half, one, or more than one REV perform better than strategies where the measurements are spatially spread out. The difference in model performance between the measurement strategies is illustrated in Figure 4 and quantified by the RMSE provided in Table 5.

[31] For the layered structure, no REV smaller than the full structure exists. This is also evident in the validation results, where the results for the layered structure are worse than for the other two structures. Only certain beneficial

combinations of measurements provide acceptable results, although there is no obvious reason why this is the case. In contrast to the periodic structure, the layered structure showed no systematic difference between parameter estimations performed with measurements taken close to each other and parameter estimations using measurements spread out over the structure. When comparing the model performance summarized in Table 5, it is obvious that all measurement strategies (Table 1) perform worse for the layered structure than for the other two structures, both in terms of average saturation and dynamic behavior. A possible exception is the connected measurement strategy, which apparently covers a very representative selection of layers. Since clearly no REV can be defined for the whole layered structure, we believe that this is a coincidence.

[32] Figure 6 (right) shows the horizontally averaged saturation profile at three times together with the simulations obtained with the best fit parameters using the all data measurement strategy. Two important points can be seen in this figure. First, the model predictions approximate the mean saturation well. However, it is evident that it is impossible

Table 5. Model Performance Determined from the Posterior Distribution Obtained with MCMC Simulation^a

Structure		All Data	Top	Connected	Spread 1	Spread 2
Random	Saturation	-6.2(-5.1)	-5.0(-3.4)	-5.3(-4.0)	-5.4(-4.1)	-5.3(-4.1)
	RCS	-6.7(-5.9)	-6.4(-4.9)	-6.4(-5.0)	-6.7(-4.9)	-6.6(-5.0)
Periodic	Saturation	-6.0(-5.0)	-5.3(-3.5)	-4.7(-3.8)	-3.5(-3.2)	-3.5(-3.2)
	RCS	-6.7(-5.9)	-6.6(-4.9)	-5.8(-4.9)	-4.1(-3.9)	-4.1(-3.9)
Layered	Saturation	-4.4(-4.0)	-2.6(-2.2)	-4.9(-3.6)	-3.5(-3.2)	-3.4(-3.3)
	RCS	-5.5(-5.2)	-3.6(-3.5)	-6.0(-4.9)	-4.0(-3.8)	-3.9(-3.7)
Full	Saturation	-4.7(-4.4)	-3.6(-2.7)	-4.2(-3.4)	-4.8(-4.1)	-4.8(-4.2)
	RCS	-4.8(-4.7)	-4.9(-4.8)	-4.8(-4.6)	-4.8(-4.7)	-4.8(-4.7)
Line A	Saturation	-4.2(-4.1)	-2.5(-2.2)	-4.3(-3.6)	-4.0(-3.6)	-4.1(-3.6)
	RCS	-4.7(-4.6)	-4.8(-4.8)	-4.8(-4.6)	-4.6(-4.4)	-4.7(-4.5)
Line B	Saturation	-4.5(-4.3)	-2.2(-2.0)	-3.7(-3.1)	-4.3(-3.7)	-4.4(-3.8)
	RCS	-4.7(-4.5)	-4.8(-4.8)	-4.5(-4.3)	-4.7(-4.4)	-4.6(-4.4)

^aNatural logarithm of minimum RMSE and mean RMSE (in parentheses) according to the definition in equation (6) for the different structures and measurement scenarios. RMSE is reported for saturation (-) and rate of change of saturations (RCS) (s⁻¹). Please note that to ease the reading of the table, the displayed values are the natural logarithms of the actual values.

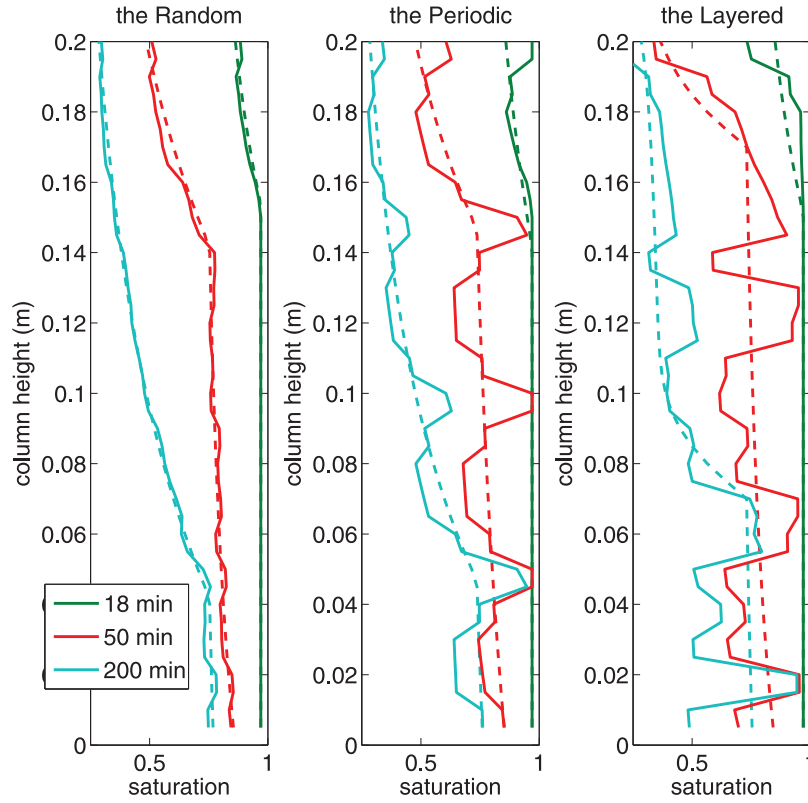


Figure 6. Example of horizontally averaged saturation profiles from measurements (solid lines) and simulations (dashed lines) for the random (left), periodic (middle), and layered (right) structure. All simulations are based on parameters estimated using the all data measurement strategy.

to match the strong fluctuations of saturation between the layers with a homogeneous model, and the effective model will always result from a compromise of the mismatches between observations and simulations in all observed layers. Second, the saturation profiles predicted with the effective model parameterizations have strong inflections near the drainage front. For example, at a height of 0.07 m and $t = 200$ min, a clear inflection can be seen for the layered structure that is not present in the random and periodic structures. This is related to a large volume of fine sand with relatively high saturation. The effective model is clearly trying to match the saturation of this particular layer, rather than the average behavior of all the layers. This leads to a poor performance of the estimated effective parameters when the average behavior is evaluated and illustrates the problem that the effective parameters can become very dependent on local structural features in heterogeneous columns.

3.2. Real Experiment Data

[33] The results for the three data sets and a range of measurement strategies for the RE case are presented in Figure 7 and Table 5. Similar to the VR case, the effective hydraulic parameters determined from the random structure with fully averaged data (full) perform well and the validation results presented in Figure 7 are reasonable for most measurement strategies. This is also confirmed by the performance criteria in Table 5. For line A and line B, which both approximate a layered structure, the validation results are more variable, and poorly performing effective hydraulic

parameters that neither match the average saturation nor the dynamics are found in some cases. Hence, the RE case also shows a need to improve effective parameter estimates, although not as clear as the layered structure in the VR case.

4. Bias Correction Using an Explicit Error Model

[34] In an attempt to obtain effective parameters with more predictive power, explicit error models are now introduced. In the context of this paper, an explicit error model is an external change of the simulated saturation values to take into account some of the soil structure present in the data without complicating the effective model and its parameters. The error models used in this paper can be understood from two different perspectives. The first perspective starts with the assumption of Gaussian distributed errors in the likelihood defined in equations (4) and (5). Equation (5) can be rewritten to explicitly consider the zero mean expectation of the distribution:

$$p(\theta|S_{\text{obs}}) \propto \exp \left[-\frac{1}{2} \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} \frac{(r_{j,i} - 0)^2}{\sigma^2} \right]. \quad (7)$$

Considering that the layers of the heterogeneous soil columns can have different volume fractions of fine and coarse sand, the assumption of a zero mean in all layers seems questionable. Instead, it would seem that allowing certain layers to deviate from a zero mean is beneficial to the

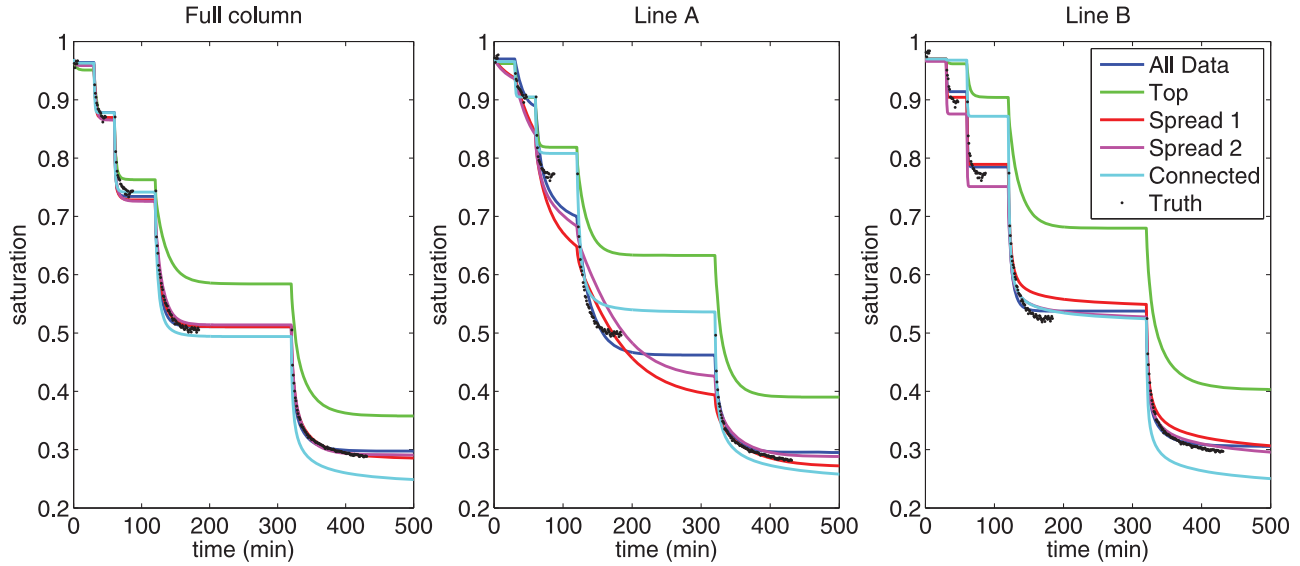


Figure 7. Validation results for the drainage scenario for different cases of the real experiment data using different measurement strategies.

search for effective parameters. An error model in the form of a nonzero mean is therefore introduced:

$$p(\theta|S_{\text{obs}}) \propto \exp \left[-\frac{1}{2} \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} \frac{(r_{j,i} - em_{j,i})^2}{\sigma^2} \right], \quad (8)$$

where em is the error model parameter.

[35] Following the work of Carter [2004], the error model of equation (8) is expanded to a fractional error model, where $em_{j,i}$ is a fraction of the simulation result:

$$em_{j,i} = S_{j,i}^{\text{org}} * e_i, \quad (9)$$

where S^{org} is the original saturation, either simulated by the flow model ($S^{\text{org}} = S_{\text{sim}}$) or observed in reality ($S^{\text{org}} = S_{\text{obs}}$) and e_i is the fractional error model parameter that is variable in space but constant in time. In this way, e_i could be considered as a bias correction in the layers.

[36] The second way of looking at the explicit error model is to see it as a transformation of the simulated (or observed) values:

$$S_{j,i}^{\text{EM}} = S_{j,i}^{\text{org}} + em_{j,i} = S_{j,i}^{\text{org}} * (1 + e_i), \quad (10)$$

where S^{EM} is the saturation after applying the error model. The probability density is then calculated as in equations (4) and (5). The form of equation (10) shows clearly that for each spatial observation ($S_{j,i}^{\text{org}}$), there will be a corresponding error model parameter (e_i).

[37] The first perspective highlights the similarity of this error model to the use of heteroscedastic error standard deviations (σ) discussed by Rigby and Stasinopoulos [1996] and the treatment of heteroscedastic non-Gaussian errors by Schoups and Vrugt [2010]. The connection between approaches that consider heteroscedastic errors and error model approaches was also pointed out by Reichert and Meileitner [2009] who noted that using an inappropriate

likelihood function and not accounting for deficiencies in the model structure can mean the same thing.

[38] The second perspective on the error model is more pragmatic and shows that the error model is related to the idea of applying external changes to the model. The similarities with the work of Kavetski *et al.* [2006], Vrugt *et al.* [2008], and Reichert and Meileitner [2009] are obvious. Kavetski *et al.* [2006] and Vrugt *et al.* [2008] allowed a change of the rainfall input to a hydrological model by introducing a rainfall multiplier for each precipitation event as an estimated parameter. Reichert and Meileitner [2009] considered both rainfall and evaporation multipliers to aid the parameter estimation. Even though the approaches have similarities, there are also clear differences. For example, the multipliers are applied to the model input, while the error model is applied to the model output. The error models used herein also have similarity with autoregressive error models that are sometimes used to improve streamflow predictions [e.g., Laloy *et al.*, 2010b]. These autoregressive models strive to reduce the impact of autocorrelated model residuals on the estimated effective model parameters by introducing an error model that assumes an error that depends on the error at the previous time step. Finally, there is also a clear connection between the error models used herein and data assimilation methods, such as the ensemble Kalman filter. Data assimilation methods estimate and update model states, hence also applying external changes as the simulation goes along, although the error model used in this study does not change in time as is the case with a Kalman filter. Ensemble Kalman filter methods have been used in reservoir modeling [e.g., Oliver and Chen, 2011], groundwater modeling [e.g., Hendricks Franssen and Kinzelbach, 2008], and have also successfully been tested in combination with global optimization for hydrological model calibration [Vrugt *et al.*, 2005].

[39] An argument against the use of an external error model could be that changing saturation outside of the flow model might violate the mass balance of the system during

drainage. Of course, the error model parameters could be constrained in such a way that mass is conserved. In this case, only redistribution of mass within the system would be allowed. Such a constraint on the mass balance is not explored in this study for the following reasons. First, the mass of the system is only truly fixed if the fluxes over the boundaries are fully controlled. In our study, boundary conditions are given as pressure heads, which means that the flux over the boundaries and the associated mass balance depend on the choice of hydraulic parameters. Second, the error model is only applied to simulations for selected measurement points that might not be representative for the whole column. In such a case, a constraint on the global mass balance might not be meaningful because a nonzero correction can be beneficial when the selected measurements are not representative for the whole column.

[40] An alternative approach to the use of error models would be to explicitly define a heterogeneous model instead of an effective homogeneous model and to subsequently estimate hydraulic parameters for each unique layer. This approach would be computationally extremely demanding except when the structure of the model is reduced to a very small number of layers. We do not follow this approach in this study, but a discussion of the benefits and drawbacks of effective homogeneous and heterogeneous modeling approaches is provided in section 4.3. The computational effort could be reduced if the different porous materials are assumed to be Miller-Miller similar [Miller and Miller, 1956]. In this case, hydraulic parameter sets of different layers could be transferred into each other using a single scaling parameter. As in the error model approach suggested herein, this would mean that a single set of hydraulic parameters and one extra parameter per layer (in this case the Miller-Miller scaling parameter) would have to be estimated. This approach is not followed here because Miller-Miller similarity is not a reasonable assumption for the sands used in the RE case [see further Vasin *et al.*, 2008]. Also, as explained further below, the calibration problem can be solved sequentially for the error model suggested here, which makes the solution fast. This would not be possible if a heterogeneity factor would be assigned to the parameters.

[41] It is important to note that even though the flow model (equation (1)) is strongly nonlinear, the error models of equations (8) and (9) are linearly dependent on the residuals (equation (4)). This means that the parameter estimation process can be performed sequentially, i.e., the error model parameters can be estimated with a simple linear estimator for each set of flow model parameters. In fact, all error model parameters in this paper can be calculated analytically. For ease of reading, only the analytical solutions are provided in the following, while the derivation is described in Appendix A.

[42] Finally, it is important to mention that the error models used in this paper are only applied to unsaturated soil since a fully saturated cell does not discriminate between the water saturation of the different materials (see Table 2). This can be done since we do not estimate S_{sat} (cf. equation (2)) and therefore have no wish to change the saturation values at full saturation. To limit the number of figures and tables in the evaluation of the error model, the only structures evaluated with the error models are the

layered structure from the VR and line A and line B from the RE.

4.1. Error Model for Virtual Reality

4.1.1. X-Parameter Error Model

[43] In the first approach, referred to as the X-parameter error model (X-EM), all seven hydraulic model parameters (Table 3) are being used together with one error model parameter per HAL. In this case, the error model is defined by

$$e_i = e_i, \quad \forall i \in \text{HAL}, \quad (11)$$

where e_i is the error model parameter for layer i from equation (9), which gives the following analytical solution for the error model parameters (for derivation, see Appendix A):

$$e_i = \frac{\sum_{j=1}^{N_i} S_{\text{obs},j,i} * S_{\text{sim},j,i}}{\sum_{j=1}^{N_i} S_{\text{sim},j,i}^2} - 1. \quad (12)$$

[44] The calculated error model parameters for the posterior distribution of the 40 HALs measurement strategy clearly show a similarity with the structure. This is shown in Figure 8, where the posterior error model parameters are shown together with a representation of the ratio between fine and coarse material in the original column, the so called structure signal. Figure 8 indicates that the X-parameter error model can handle the problem with the large differences in saturation between the layers due to the structure quite well, as evidenced from the high (absolute) error model

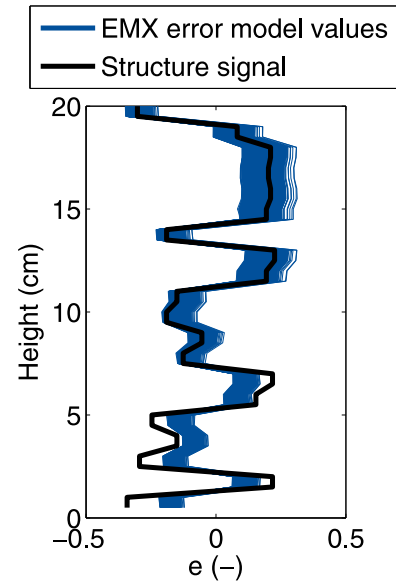


Figure 8. Comparison between the error model parameter values of the X-parameter error model, computed from the posterior parameter distribution of the all data measurement strategy, and the structure signal used in section 4.1.2. Please note that the scale on the x axis refers to the error model values and that the structure signal is superimposed to show the similarities in shape. It is not to scale.

parameter values that are associated with layers that show a strong deviation from the mean of the structure signal. The performance of the X-EM in the validation scenarios is shown in Figure 9, where the average saturation over time is shown for the different measurement strategies (Table 1). Figure 10 shows the range of RMSE values for saturation and change in saturation (equation (6)) determined from effective hydraulic parameters drawn from the posterior distribution for the X-EM approach.

[45] As can be seen in Figure 9, the use of the error model helps to improve model performance for most measurement strategies. Only the performance of the connected measurement strategy strongly decreased with a particularly strong offset in the average saturation after implementation of the error models. This indicates that a simple error model that only depends on the simulations and observations can be useful to improve model performance of poorly performing models. However, the introduction of an error model might also entail a deterioration of predictive power because of a lack of restrictions on the error model parameters.

4.1.2. Two-Parameter Error Model

[46] To add a forced dependence of the error model on the structure, a two-parameter error model that requires prior knowledge of the structure is tested:

$$e_i = Z_{\text{str},i} * e_{\text{shape}} + e_{\text{bias}}, \quad \forall i \in \text{HAL}, \quad (13)$$

where Z_{str} is a structure signal containing prior information about the structure, e_{shape} is an error model parameter that changes the shape of Z_{str} , and e_{bias} is a second error model parameter that corrects for offsets in Z_{str} . The structure signal is prescribed, meaning that only two error model parameters are estimated. When the $\sum_{j=1}^{N_t} \sum_{i=1}^{N_s}$ is denoted as $\sum \sum$, the analytical solution to equation (13) is (for derivation, see Appendix A)

$$\begin{pmatrix} \sum \sum (S_{\text{sim},j,i}^2 Z_{\text{str},i}^2) & \sum \sum (2S_{\text{sim},j,i}^2 Z_{\text{str},i}) \\ \sum \sum (2S_{\text{sim},j,i}^2 Z_{\text{str},i}) & \sum \sum (S_{\text{sim},j,i}^2) \end{pmatrix} \begin{pmatrix} e_{\text{shape}} \\ e_{\text{bias}} \end{pmatrix} = \begin{pmatrix} \sum \sum (S_{\text{obs},j,i} S_{\text{sim},j,i} Z_{\text{str},i} - S_{\text{sim},j,i}^2 Z_{\text{str},i}) \\ \sum \sum (S_{\text{obs},j,i} S_{\text{sim},j,i} - S_{\text{sim},j,i}^2) \end{pmatrix}. \quad (14)$$

In contrast to the previous approach, these error model parameters are stationary in both time and space. The choice of the structure signal Z_{str} is crucial to the success of this approach. If a good estimate of the volume fraction of the materials in the different layers exists, a structure signal of the following form can be used:

$$Z_{\text{str},i} = \text{ratio}_{\text{fine},i}, \quad \forall i \in \text{HAL}, \quad (15)$$

where $\text{ratio}_{\text{fine},i}$ is the volume fraction of fine material in layer i . The shape of Z_{str} for the layered structure is shown in Figure 8. The results of the validation scenarios are shown in Figures 10 and 11. An improvement of the model performance by the two-parameter error model is apparent in most cases. Again, only the predictive power of the connected measurement strategy (see Table 1) deteriorates. The differences between measured and modeled saturation are, however, smaller than for the previous error model. Similar to the X-parameter error model, the top measurement strategy also showed considerable deviations between measured and modeled saturation. For this measurement strategy, no difference in the performance can be seen between the X-parameter error model and the two-parameter error model. It should, however, be noted that both error model approaches clearly showed improved model performance in the validation for the top measurement strategy (Figures 9 and 11). This illustrates the necessity for validation data to properly assess the added value of any error model.

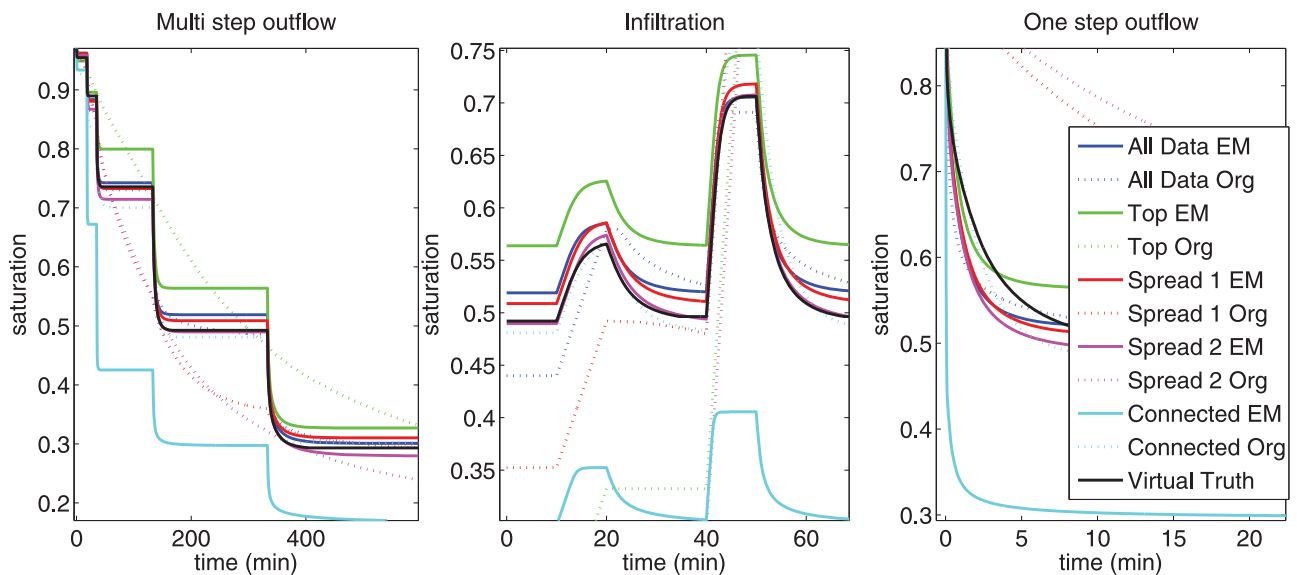


Figure 9. Results of the validation scenarios for the layered structure. Comparison between results obtained with and without the use of the X-parameter error model for different measurement strategies.

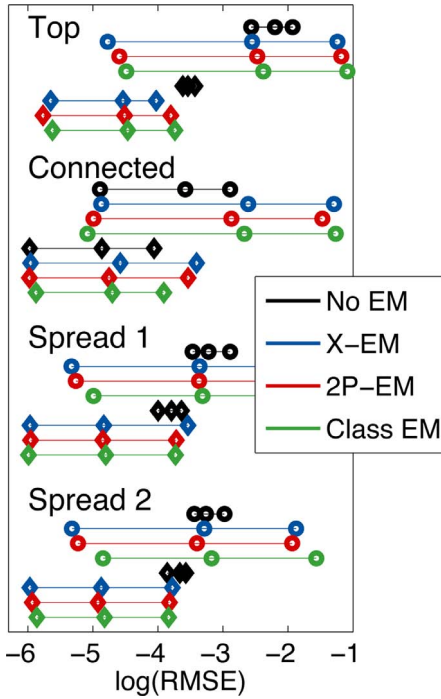


Figure 10. Range of the RMSE derived from the posterior parameter distribution for the layered structure. Comparison between RMSE ranges obtained with and without the use of several error model approaches. The markers show minimum (left marker), mean (middle marker), and maximum (right marker) RMSE (equation (6)). Dots show RMSE for saturation (–) and diamonds show RMSE for the rate of change of saturation (s^{-1}).

[47] Obviously, the two-parameter error model requires considerable prior information since a representative description of the structure in the column is needed and the use of an inappropriate structure could easily yield inappropriate

results. Such cases have been tested, but are not presented here. Therefore, the two-parameter error model is not likely to be useful outside of any virtual reality or strictly controlled laboratory experiment.

4.1.3. Class Based Error Model

[48] In practice, it might be difficult to obtain detailed information about subsurface structure. However, a rough idea about structural properties and materials might exist. A more realistic structure signal could then be a signal that is simply divided into n classes, where each class has an error parameter of its own:

$$e_i = \begin{cases} e_1 & \text{if } i \in \text{class 1} \\ e_2 & \text{if } i \in \text{class 2} \\ \dots & \\ e_n & \text{if } i \in \text{class } n \end{cases}, \quad \forall i \in \text{HAL} \quad (16)$$

which results in the following analytical solution, here shown for class n :

$$e_n = \frac{\sum_{j=1}^{N_i} \sum_{i \in \text{class } n} S_{\text{obs},j,i} * S_{\text{sim},j,i}}{\sum_{j=1}^{N_i} \sum_{i \in \text{class } n} S_{\text{sim},j,i}^2} - 1. \quad (17)$$

If n is the same as the number of layers, this model is identical to the X-parameter error model (equation (11)). The error model is tested here for the case of $n = 3$ for the layered structure. Once the number of classes is fixed, the layers have to be assigned to the classes. Figure 12 together with Figure 10 shows the results for the validations scenarios using $n = 3$ and classes based on the fine sand fraction. Class boundaries were chosen such that each class includes a similar number of layers (less than 48% fine sand, 48%–93% fine sand, and more than 93% fine sand).

[49] The class based error model on average performs worse than the other two error models in the validation

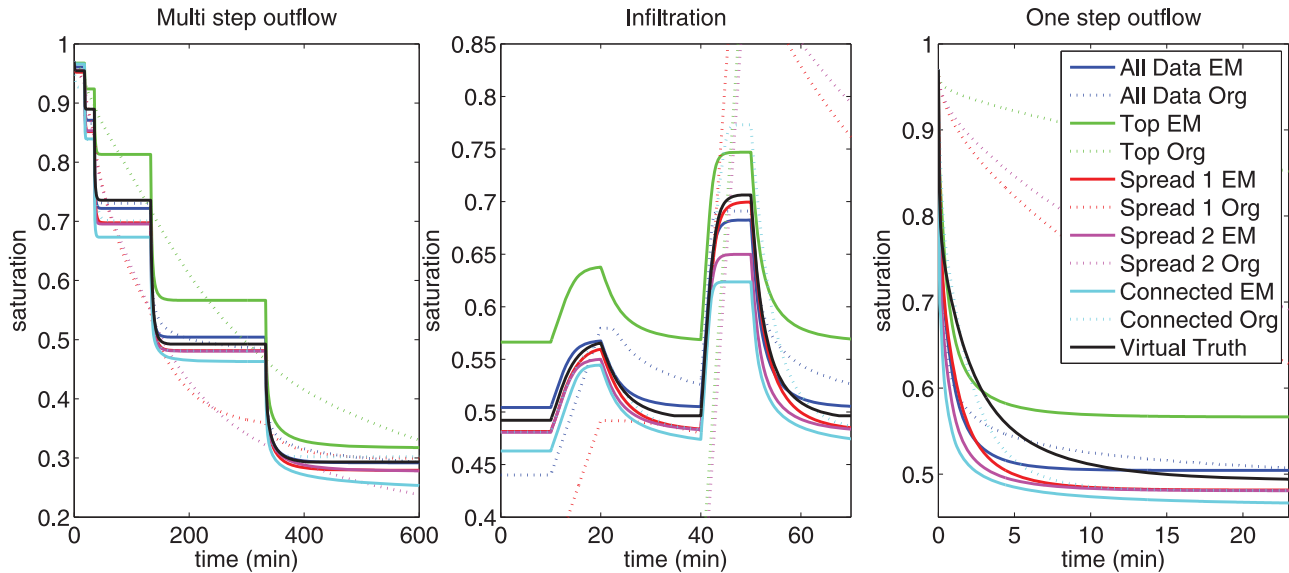


Figure 11. Results for the validation scenarios for the layered structure. Comparison between results obtained with and without the use of the two-parameter error model for different measurement strategies.

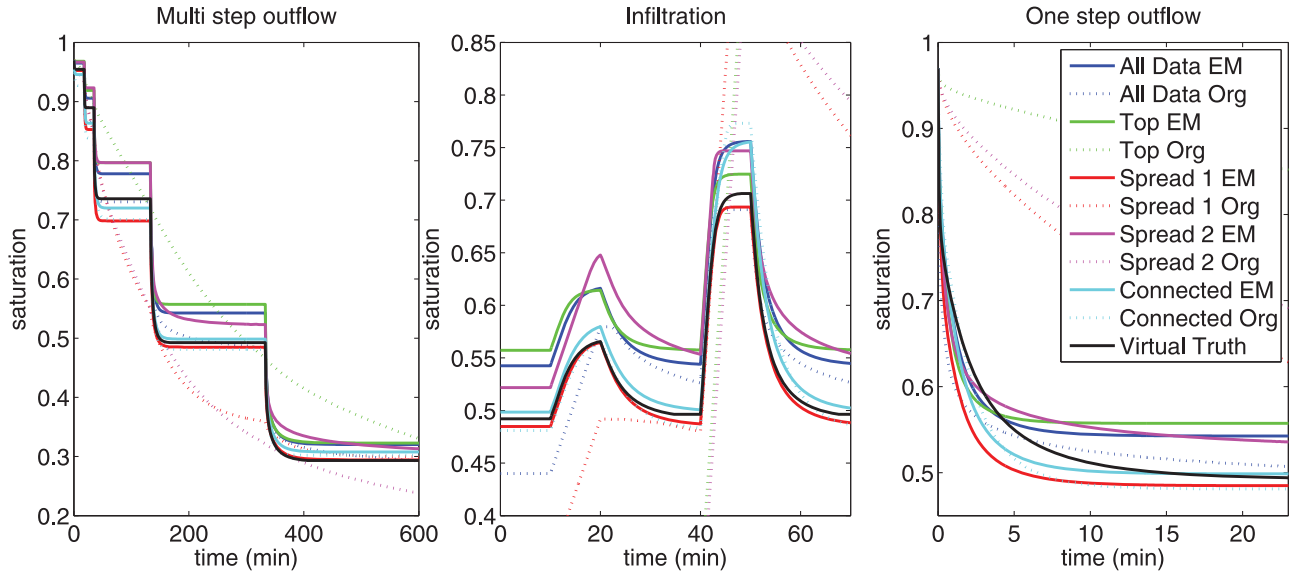


Figure 12. Results for the validation scenarios for the layered structure. Comparison between results obtained with and without the use of class based error model with $n = 3$ classes for different measurement strategies.

scenarios (Figure 12). A similar behavior as with the two-parameter error model can be seen, but the positive effects of the two-parameter error model are more evident. This is logical since the class based model is a coarse approximation of the two-parameter model. Compared with the X-parameter error model, most measurement strategies perform worse. However, the restriction of the number of classes also avoids a strong deterioration of predictive power between estimation and validation for all measurement strategies and even the connected measurement strategy performed reasonably well.

[50] As with the two-parameter error model, the difficulty with this method is the selection of the structure signal. However, it is not unreasonable to assume a certain amount of knowledge about the structure and even the saturation data can provide information on layering. Of course, the selection of an inappropriate structure has a negative impact on the results (not shown).

4.2. Real Experiment Data

[51] The three error models are evaluated for the real data cases line A and line B using the connected and spread 1 measurement strategies (Table 1). Validation results are shown for line A in Figures 13 and 14 and for line B in Figures 13 and 15. As one can expect, the results are less clear when dealing with real data. The figures indicate that average saturation improves in some cases with the introduction of the error model, but in other cases the performance decreases. The decrease in model performance is strongest for cases where the original calibration agreed well with the validations. This is no surprise since, as pointed out before, the error models may violate the mass balance. The effect of this is more evident when using the RE data since other error sources besides missing structure affect parameter estimation and may lead to a decrease in performance in the validation. Interestingly, the use of error models did not have any negative effect on the dynamics of the average saturation in the column. On the contrary, the dynamic behavior in the validation improved in most cases. The best

example of this can be seen in the right plot of Figure 14, where the dynamic improvements are obvious, but especially the two-parameter error model shows a strong deviation in average saturation. Again, the modeler needs to

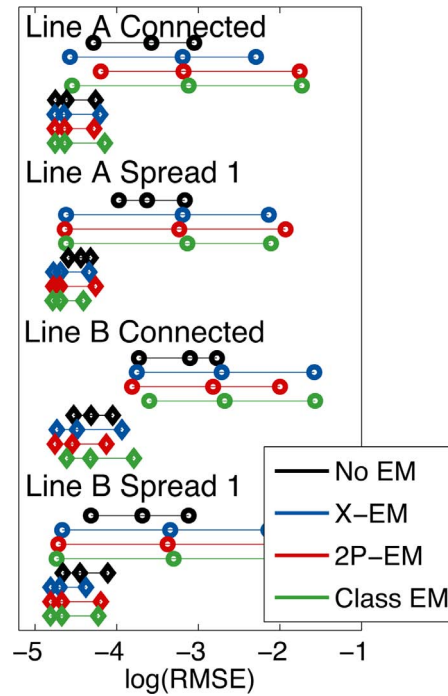


Figure 13. Range of the RMSE derived from the posterior parameter distribution for the RE case. Comparison between RMSE ranges obtained with and without the use of several error model approaches. The markers show minimum (left marker), mean (middle marker), and maximum (right marker) RMSE (equation (6)). Dots show RMSE for saturation (–) and diamonds show RMSE for the rate of change of saturation (s^{-1}).

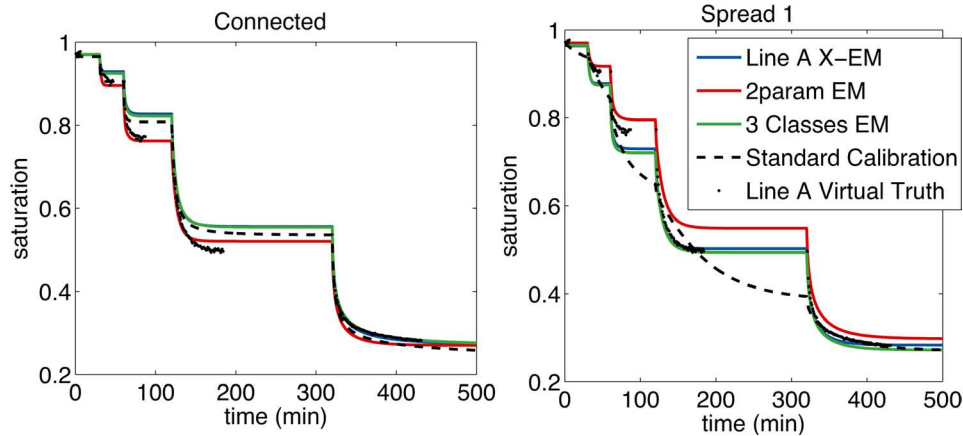


Figure 14. Results of the validation scenario for line A in the RE using the connected and the spread 1 measurement strategies. Comparison between different error model approaches.

weight the performance criteria and decide what aspects of the model simulations need to be reproduced reliably.

[52] Interestingly, the two-parameter error model does not seem to perform better than the class error model and in many cases the results are worse. This difference to the VR experiments can probably be attributed to the structure signal used in the RE. As mentioned before, imperfect packing and interfaces between the different inclusions in the RE are two possible reasons why the used signal may not be ideal. In general, the results for the real data confirm the findings for the virtual reality data. This suggests that the use of an external error model could be a useful strategy to estimate effective model parameters with more predictive ability for a strongly heterogeneous reality.

4.3. Error Model Summary and Discussion

[53] It comes as no surprise that many of the basic parameter estimation scenarios presented in section 3 showed poor results. Fitting an effective homogeneous model to a strongly heterogeneous reality is fundamentally problematic. Since we aimed to reproduce the averaged water fluxes and water contents, the validation data used in this study

focused on average saturation. The internal water distribution was not considered and was, most likely, not well reproduced. The performance criteria used to summarize the model performance during validation focused on the reproduction of the average saturation and the flow dynamics. The importance assigned to these performance criteria depends on the aim of the modeling. Our results show that an improvement for one performance criterion can be associated with deterioration in another criterion.

[54] Some of the issues associated with the estimation of effective hydraulic parameters could have been avoided by the use of a detailed heterogeneous model. If layers with individual model parameters would be used in such a model, the model predictions could improve substantially [e.g., Durner *et al.*, 2008] and the resulting model would probably be a better physical representation of the reality. Furthermore, such a layered model could probably better represent the internal water distribution, which cannot be expected for effective homogeneous models with or without an additional error model. The drawbacks of using a heterogeneous model are also evident. First, boundaries between different layers would need to be defined, which is

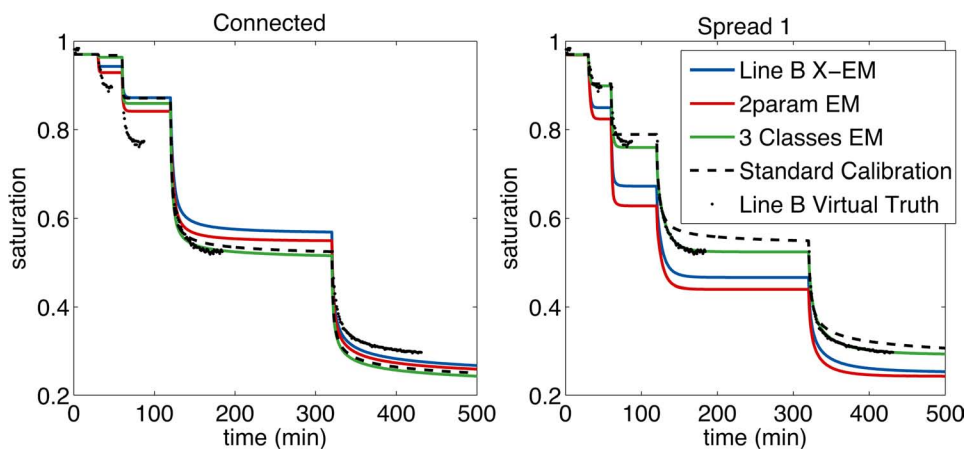


Figure 15. Results of the validation scenario for line B in the RE using the connected and the spread 1 measurement strategies. Comparison between different error model approaches.

difficult when limited information is available (i.e., when the measurements do not cover the entire domain). To set up such a heterogeneous model, one would need a more complete idea of the thickness of the different layers than required for the error model approaches. Second, there are many more parameters to estimate, and in contrast to the error model approaches, they cannot be estimated sequentially. This could lead to far larger time requirements for parameter estimation, unless prior knowledge about parameter values is explicitly added to the parameter estimation process as prior information. Finally, the resulting parameters become dependent on the position of the observations in the column and the subsections of data used. It should be noted, once again, that if one has this prior information to set up a layered model (for example as we do in our all data case), this would surely be a more suitable approach. The error model approaches discussed herein are designed for cases when only smaller subsections of data are available and prior information to set up a layered model is not available.

[55] In an attempt to account for unresolved structure in the estimation of effective model parameters, three different error models that require different amounts of prior information were considered for bias correction. The highest level of prior knowledge is required by the two-parameter error model, which needs a representative description of the structure in the column. This error model performed well and improved results for most of the VR data. It also showed a clear improvement in model performance for the dynamics of the RE data.

[56] In contrast, the most flexible approach is the X-parameter error model that does not require any prior information at all. With this error model, most predictions of average saturation improved. However, the X-parameter error model also gave the maximum likelihood parameter set (best calibration result) that performed worst in the validation scenarios (connected measurement strategy in Figure 9). This suggests that one should be careful when applying such a flexible error model, especially when the effective hydraulic parameters estimated without error model perform well during validation. The X-parameter error model approach also has a large need for high quality data since the full error model in its analytical form is only based on the measurements and simulations and no other information.

[57] The third error model was the class error model approach, which is a compromise between the previous two. It contains a rough estimate of structure, but the relation between the classes is not fixed a priori as in the two-parameter case. In this case, the use of the error model resulted in a clear improvement of the validation performance for the flow dynamics, which improved or conserved prediction quality for all tested cases.

[58] When comparing the spread of the RMSE between different error models as caused by the posterior uncertainty in the estimated hydraulic parameters for the multi-step outflow validation scenario (shown in Figure 10), some observations can be made. First, the use of error models does not lead to higher RMSE values when comparing calibrations with and without error model, and in most cases an improvement in the minimum RMSE is observed. This implies that each posterior distribution contains parameter sets that also perform well for the validation scenarios, although this parameter set is not necessarily the

same as the one that has the maximum likelihood for the calibration data. Indeed, it is obvious from comparing the figures that the most likely parameter sets for the calibration data are not the parameter sets that provide the best predictions in the validation scenarios (e.g., compare Figure 9 with Figure 10). This further highlights the need for validation data to identify these parameter sets. Second, it is evident that the range of RMSE values increases dramatically when an error model is introduced. This is especially true for cases that perform poorly without an error model and not so much for the connected measurement strategy that performs well without error model (see Table 1). This increases our belief that the effective parameter ranges obtained with the error models are more representative for the system.

[59] The range of RMSE values derived from the posterior distribution should, however, be interpreted with care. The standard deviation of the residuals was assumed to be high ($\sigma = 0.15$, cf. equation (8)) to reflect the large variation of the saturation seen in the data, especially in the case of the periodic and the layered structure. Smaller values of σ were also tested and resulted in similar performances of the model with the best suggested parameters but lower acceptance rates and hence longer run times for the MCMC algorithm. In the VR case, σ should reflect errors introduced by spatial averaging of the observations and errors associated with unresolved structure. When the error model is introduced, it is expected that some of the errors introduced by unresolved spatial structure are compensated, which suggests that σ should be lower. Nevertheless, we decided to keep the same standard deviation when introducing the error models, leading to a potential overdispersion of the posterior distribution. The same is true for RE case, although additional sources of error make the choice of an appropriate value for σ even more difficult. As pointed out before, the primary interest of this work was the predictive ability of the maximum likelihood estimates of the effective hydraulic parameters, and the posterior uncertainty of the model parameters and predictions was of secondary importance. We only report how ranges of RMSE values are affected by the introduction of the error models. Indeed, to appropriately determine predictive uncertainty a different approach should be used to determine σ . Recent studies have explored the estimation of σ as an additional model parameter in a generalized likelihood function [Schoups and Vrugt, 2010]. Although this method to determine a more appropriate value of σ is promising, it is beyond the scope of this paper.

[60] A relevant question to address when using MCMC simulations is the appropriateness of the likelihood function. With the Gaussian assumption of equation (7), the residuals should be unimodal, normally distributed and centered on zero. For the random structure as well as for the measurement strategies that performed well for the periodic structure, this is the case (results not shown here). For the layered structure, clear multimodality is often seen. After the introduction of the error models, this multimodality decreases and a more Gaussian shape of the residuals is obtained. An example of this effect is shown in Figure 16.

[61] It is interesting to note that the use of an error model improved the predictive power of the estimated parameters when only a subset of observations was used. It has been

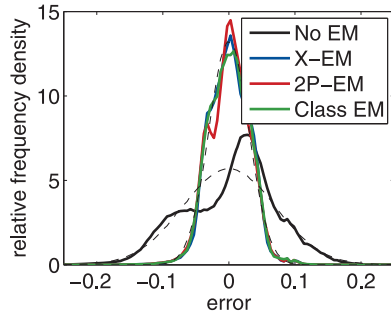


Figure 16. Histogram of the residuals for the measurement strategy spread 2 for the layered structure with and without error models. Dashed lines show two superimposed normal distributions for comparison.

discussed that hydraulic parameters estimated from outflow data cannot represent the internal water distribution [Bayer *et al.*, 2005; Durner *et al.*, 2008; Laloy *et al.*, 2010a]. One could therefore expect that model parameters estimated with only a few internal observations should not be able to represent the outflow. In this case, the error model clearly helped to bridge the gap between the internal states and the resulting outflow.

[62] One of the reasons why the use of error models provided flow models with improved predictive power is because of the high information content in the data due to horizontal spatial averaging. In this study we relied on radiography data obtained in the laboratory. Similar data could be obtained in the field with borehole ground penetrating radar (GPR) as reviewed by Huisman *et al.* [2003]. Especially the zero offset profile mode that provides soil water content profiles with a high enough temporal resolution to capture transient processes is promising for field applications. Emerging within the field of advanced inverse methods are also coupled inversion approaches, such as Hinnell *et al.* [2010], that combine geophysical and forward flow modeling in one parameter estimation framework. It could be tested in future work if the coupled inversion of zero offset profile GPR data with an effective flow model would profit from introducing an external error model.

5. Conclusions

[63] The aim of this study was to test external error models for estimating hydraulic parameters of unsaturated flow models for the case that an effective homogeneous model is fitted to observations with small observation volumes in heterogeneous media. The tests of different external error models demonstrate that in cases where observation volumes do not cover or nearly cover an REV, such as the layered column in this study, the use of any of the suggested error models can improve the performance of the effective homogeneous model for the validation scenarios. It is also demonstrated that if prior knowledge of the soil structure is available to set up a two-parameter or a class-based error model, using a limited amount of observation from a small subsection of the column can provide model predictions with reasonable to good performance in the validation scenarios. This suggests that if a limited amount of observations is available, parameter estimation results can still be

acceptable if knowledge of soil structure is available. Hence an external error model can be a useful approach if no REV of the medium can be defined or when an effective model is sought for a larger scale than the observations.

Appendix A: Analytical Derivation of the Error Models

[64] Combining equations (4), (8), and (9), the posterior distribution $p(\theta|S_{\text{obs}})$ can be written as

$$p(\theta|S_{\text{obs}}) \propto \exp \left\{ -\frac{1}{2} \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} \frac{[S_{\text{obs},j,i} - S_{\text{sim},j,i}(1 + e_i)]^2}{\sigma^2} \right\}.$$

If we are only interested in the optimum of this distribution, we can rewrite the posterior sampling problem as a minimization problem:

$$f = \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} [S_{\text{obs},j,i} - S_{\text{sim},j,i}(1 + e_i)]^2.$$

For the X-parameter error model of equation (10), this means finding the optimal values of e_i for each layer i . The previous equation then becomes

$$f_i = \sum_{j=1}^{N_t} \{ [S_{\text{obs},j,i} - S_{\text{sim},j,i}(1 + e_i)]^2 \} \quad \text{with} \quad \sum_i (f_i) = f.$$

Given fixed observations (S_{obs}) and simulations (S_{sim}), the e_i that minimizes the function f_i can be found by equating the derivative of f_i with respect to e_i to zero:

$$\frac{\partial f_i}{\partial e_i} = 0 \iff \sum_{j=1}^{N_t} \{ -2S_{\text{sim},j,i} [S_{\text{obs},j,i} - S_{\text{sim},j,i}(1 + e_i)] \} = 0$$

and to solve for e_i :

$$e_i = \frac{\sum_{j=1}^{N_t} [S_{\text{sim},j,i}(S_{\text{obs},j,i} - S_{\text{sim},j,i})]}{\sum_{j=1}^{N_t} (S_{\text{sim},j,i}^2)} = \frac{\sum_{j=1}^{N_t} (S_{\text{sim},j,i} S_{\text{obs},j,i})}{\sum_{j=1}^{N_t} (S_{\text{sim},j,i}^2)} - 1.$$

The expansion of the previous equation to the class based error model of equation (16) is straightforward and results in

$$e_n = \frac{\sum_{j=1}^{N_t} \sum_{i \in \text{class } n} (S_{\text{obs},j,i} * S_{\text{sim},j,i})}{\sum_{j=1}^{N_t} \sum_{i \in \text{class } n} (S_{\text{sim},j,i}^2)} - 1.$$

[65] The analytical solution of the two-parameter error model (equation (13)) can be derived by minimizing the function g :

$$g = \sum_{j=1}^{N_t} \sum_{i=1}^{N_s} [S_{\text{obs},j,i} - S_{\text{sim},j,i}(1 + Z_{\text{str},i}e_{\text{shape}} + e_{\text{bias}})]^2.$$

Given fixed observations (S_{obs}), simulations (S_{sim}), and structure signal (Z_{str}), the values of e_s and e_b that minimize

g can be found by equating the derivative of g with respect to e_b and e_s to zero:

$$\frac{\partial g}{\partial e_s} = 0,$$

$$\frac{\partial g}{\partial e_b} = 0.$$

Solving the system of equations for e_b and e_s and rewriting in matrix-vector form results in

$$\begin{pmatrix} e_s \\ e_b \end{pmatrix} = \begin{pmatrix} \sum \sum (S_{\text{sim},j,i}^2 Z_{\text{str},i}^2) & \sum \sum (S_{\text{sim},j,i}^2 Z_{\text{str},i}) \\ \sum \sum (S_{\text{sim},j,i}^2 Z_{\text{str},i}) & \sum \sum (S_{\text{sim},j,i}^2) \end{pmatrix}^{-1} \begin{pmatrix} \sum \sum (S_{\text{obs},j,i} S_{\text{sim},j,i} Z_{\text{str},i} - S_{\text{sim},j,i}^2 Z_{\text{str},i}) \\ \sum \sum (S_{\text{obs},j,i} S_{\text{sim},j,i} - S_{\text{sim},j,i}^2) \end{pmatrix}.$$

Notation

EM	error model.
HAL	horizontally averaged layer.
MCMC	Markov chain Monte Carlo.
MSO	multistep outflow (experiment).
ORG	original (no EM) parameter estimation.
VR	virtual reality.
RE	real experiment.
RMSE	root mean squared error.

[66] **Acknowledgments.** We would like to thank the Associate Editor Jasper Vrugt and the three anonymous reviewers for their comprehensive reviews and constructive comments. We also acknowledge the RRZN Hannover for facilitating computation power.

References

- Ashby, S. F., and R. D. Falgout (1996), A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations, *Nucl. Sci. Eng.*, **124**, 145–159.
- Bayer, A., H. J. Vogel, O. Ippisch, and K. Roth (2005), Do effective properties for unsaturated weakly layered porous media exist? An experimental study, *Hydrol. Earth Syst. Sci.*, **9**, 517–522.
- Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology, *J. Hydrol.*, **249**, 11–29.
- Brooks, R., and A. Corey (1966), Properties of porous media affecting fluid flow, *J. Irrig. Drain. Eng.*, **92**, 61–88.
- Carter, J. N. (2004), Using Bayesian statistics to capture the effects of modelling errors in inverse problems, *Math. Geol.*, **36**, 187–216.
- Coppola, A., A. Basile, A. Comegna, and N. Lamaddalena (2008), Monte Carlo analysis of field water flow comparing uni- and bimodal effective hydraulic parameters for structured soil, *J. Contam. Hydrol.*, **104**, 153–165, doi:10.1016/j.jconhyd.2008.09.007.
- Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, **46**, W05525, doi:10.1029/2009WR008377.
- Durner, W. (1994), Hydraulic conductivity estimation for soils with heterogeneous pore structure, *Water Resour. Res.*, **30**, 211–233.
- Durner, W., U. Jansen, and S. C. Iden (2008), Effective hydraulic properties of layered soils at the lysimeter scale determined by inverse modelling, *Eur. J. Soil Science*, **59**, 114–124, doi:10.1111/j.1365-2389.2007.00972.x.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman and Hall/CRC, Boca Raton, FL.
- Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, **44**, W09408, doi:10.1029/2007WR006505.
- Hinnell, A. C., T. P. A. Ferré, J. A. Vrugt, J. A. Huisman, S. Moysey, J. Rings, and M. B. Kowalsky (2010), Improved extraction of hydrological information from geophysical data through coupled hydrogeophysical inversion, *Water Resour. Res.*, **46**, W00D40, doi:10.1029/2008WR007060.
- Huisman, J. A., S. S. Hubbard, J. D. Redman, and A. P. Annan (2003), Measuring soil water content with ground penetrating radar: A review, *Vadose Zone J.*, **2**, 476–491.
- Jones, J. E., and C. S. Woodward (2001), Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems, *Adv. Water Resour.*, **24**, 763–774.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, **42**, W03408, doi:10.1029/2005WR004376.
- Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Statist. Soc. Ser. B.*, **63**, 425–464.
- Kollet, S. J., and R. M. Maxwell (2005), Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, *Adv. Water Resour.*, **29**, 945–958, doi:10.1016/j.advwatres.2005.08.006.
- Kumar, S., M. Sekhar, D. V. Reddy, and M. S. M. Kumar (2010), Estimation of soil hydraulic properties and their uncertainty: Comparison between laboratory and field experiment, *Hydrol. Processes*, **24**, 3426–3435, doi:10.1002/hyp.7775.
- Laloy, E., M. Weynants, C. L. Biielders, M. Vanclooster, and M. Javaux (2010a), How efficient are one-dimensional models to reproduce the hydromechanic behavior of structured soils subject to multi-step outflow experiments?, *J. Hydrol.*, **393**, 37–52, doi:10.1016/j.jhydrol.2010.02.017.
- Laloy, E., D. Fasbender, and C. L. Biielders (2010b), Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach, *J. Hydrol.*, **380**, 82–93, doi:10.1016/j.jhydrol.2009.10.025.
- Mertens, J., H. Madsen, M. Kristensen, D. Jacques, and J. Feyen (2005), Sensitivity of soil parameters in unsaturated zone modelling and the relation between effective, laboratory and in situ estimates, *Hydrol. Processes*, **19**, 1611–1633, doi:10.1002/hyp.5591.
- Miller, E. E., and R. D. Miller (1956), Physical theory for capillary flow phenomena, *J. Appl. Phys.*, **27**, 324–332.
- Oliver, D. S., and Y. Chen (2011), Recent progress on reservoir history matching: A review, *Comp. Geosci.*, **15**, 185–221, doi:10.1007/s10596-010-9194-2.
- Papafotiou, A., et al. (2008), From the pore scale to the lab scale: 3-D lab experiment and numerical simulations of drainage in heterogeneous porous media, *Adv. Water Res.*, **31**, 1253–1268.
- Reichert, P., and J. Meileitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time dependent parameters, *Water Resour. Res.*, **45**, W10402, doi:10.1029/2009WR007814.
- Rigby, R. A., and D. M. Stasinopoulos (1996), A semi-parametric additive model for variance heterogeneity, *Stat. Comput.*, **6**, 57–65.
- Schelle, H., S. C. Iden, A. Peters, and W. Durner (2010), Analysis of the agreement of soil hydraulic properties obtained from multistep-outflow and evaporation methods, *Vadose Zone J.*, **9**, 1080–1091, doi:10.2136/vz2010.0050.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, **46**, W10531, doi:10.1029/2009WR008933.
- Valiantzas, J. D., and P. A. Londra (2008), Direct determination of the Brooks-Corey hydraulic functions by fitting an extended power function to the outflow method data, *J. Hydrol.*, **362**, 128–133, doi:10.1016/j.jhydrol.2008.08.021.
- Vasin, M., P. Lehmann, A. Kaestner, R. Hassanein, W. Nowak, R. Helmig, and I. Neuweiler (2008), Drainage in heterogeneous sand columns with different geometric structures, *Adv. Water Resour.*, **31**, 1205–1220.
- Vereecken, H., R. Kasteel, J. Vanderborght, and T. Harter (2007), Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review, *Vadose Zone J.*, **6**, 1–28, doi:10.2136/vzj2006.0055.
- Vereecken, H., J. A. Huisman, H. Bogaen, 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, doi:10.1029/2008WR006829.
- Vogel, H. J., and K. Roth (2003), Moving through scales of flow and transport in soil, *J. Hydrol.*, **272**, 95–106.

- Vogel, H. J., A. Samouëlian, and O. Ippisch (2008), Reprint of “multi-step and two-step experiments in heterogeneous porous media to evaluate the relevance of dynamic effects,” *Adv. Water Resour.*, 31, I–VIII, doi:10.1016/j.advwatres.2007.08.013.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39(8), 1201, doi:10.1029/2002WR001642.
- 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 strength of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Zhang, D. (2002), *Stochastic Methods for Flow in Porous Media*, Academic, San Diego.
- Zurmühl, T., and W. Durner (1998), Determination of parameters for bimodal hydraulic functions by inverse modeling, *Soil Sci. Soc. Am. J.*, 62, 874–880.