ON THE IDENTIFICATION OF SOIL HYDRAULIC PARAMETERS

N. Fajraoui[*], D. Calogine[†], F. Lehmann[*],T.A. Mara[†] and A. Younes[*]

[*] Laboratoire d'Hydrologie et de Géochimie de Strasbourg
Université de Strasbourg, CNRS/UMR7517
1 rue Blessig, 67084 Strasbourg, France

[†] PIMENT
Université de La Réunion,
15 Avenue René Cassin, 97715 Saint-Denis, Réunion

e-mail: Fajraoui@unistra.fr

Key words: saturated-unsaturated flow, polynomial chaos expansion, MCMC

Summary. Drainage experiments are commonly conducted to determine soil hydraulic parameters. Experiments generally measure the temporal evolution of the cumulative outflow as well as pressure heads and water contents at several levels in the column. Then the analysts try to calibrate the model by fitting the model response to the data series. The aim of the present work is to investigate the quality of inverting Richards-Van Genuchten Mualem (R-VGM) equations prior to experiment. For this purpose, we invert R-VGM in a Bayesian framework from artificial data series. Prior to the inverse modeling, global sensitivity analyses of R-VGM are performed with polynomial chaos expansion method. Then, we use Monte Carlo Markov Chain (MCMC) to estimate the marginal posterior distribution of the parameters. Such an approach allows to address the following issues: Is the model overparametrized? What are the parameters that can reasonably be estimated? And from which measurements?

1 PROBLEM STATEMENT

The simulation of saturated-unsaturated water flow in a vertical soil column is achieved with the one dimensional pressure-head based form of the Richards equation (RE),

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

where h is the pressure head. $c(h)$ is the specific moisture capacity $\left(c(h)=d\theta/dh\right)$, $K\left(h\right)$ is the hydraulic conductivity.

The interdependencies of pressure head, hydraulic conductivity and water content are characterized by constitutive equations. In this work, we use the standard model of Mualem (1976)[1]-van Genuchten (1980)[2]:

$$S_e(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \begin{cases} \dfrac{1}{\left(1 + |\alpha h|^n\right)^{(1-1/n)}} & h < 0 \\ 1 & h \geq 0 \end{cases} \tag{2}$$

$$K(S_e) = K_s S_e^{1/2} \left[1 - \left(1 - S_e^{(n/(n-1))}\right)^{(1-1/n)} \right]^2$$

where S_e is the effective saturation, θ_r the residual water content, θ and θ_s are respectively the volumetric and saturated water content, K_s the saturated conductivity, α, and n the so-called van Genuchten soil parameters.

Although the mixed form of the RE presents better conservation properties, it was demonstrated that higher order numerical integration is effective for solving the pressure head form (2) [3,4,5]. For the spatial discretization of (1), we use a finite volume method with a uniform mesh discretization of 80 cells for the whole domain (sand + plate). The time discretization is performed with higher order methods via the Method Of Lines (MOL). The advantage of the MOL is that error checking, robustness, order selection and time step adaptativity features available in sophisticated ODE/DAE codes can be applied to the time integration of the PDE. In this study, the parameters $K_s, K_p, \lambda, \theta_r, \alpha$ and n will be subject of calibration using synthetic measurements of pressure and cumulative outflow. (Table1).

Parameter	Values used for generation of synthetic data sets	min	max
K_s	7.07	3	10.0
K_p	0.308	0.1	0.5
λ	0.5	0.1	2.0
θ_r	0.0575	0.0	0.1
α	0.08	0.01	0.15
n	2.68	1.5	9

Table 1 : Example of the construction of one table

For use in the inverse analysis, synthetic time series of cumulative outflow and pressure head at depth 4 cm from the top were disturbed by normally distributed random error with mean value of 0 and standard deviations of 0.01 and 0.1 respectively.

2 POLYNOMIAL CHAOS EXPANSION

Let us consider a physical model represented by the mathematical model $M(\Theta)$ at a given space-time location where $\Theta = \{\theta_1,......,\theta_M\}$ are independent random parameter. Its development in polynomial chaos expansion of order not exceeding p is the following:

$$M(\Theta) = \sum_{|\alpha| \leq p} M_\alpha \psi_\alpha (\theta_1,...,\theta_M) \qquad (3)$$

where $|\alpha| = \sum_{i=1}^{M} \alpha_i$, ψ_α is the multidimensional orthogonal polynomial. Legendre orthogonal polynomials are used in this work since a uniform distribution is assumed for each random parameter. M_α are the PC coefficients. The number of polynomial coefficients is $P+1 = (M+p)!/(M!p!)$. To compute the polynomial coefficients different sampling techniques are proposed in the literature. Tatang et al. (1998)[6] proposed the full probabilistic collocation while Sudret (2008)[7] employed an incomplete probabilistic collocation design. In this work, we use Quasi Monte Carlo (QMC) sequence. It has been shown that the QMC performs better in terms of convergence property than the other sampling techniques because it allows to uniformly cover the parameter space. This has the advantage of highlighting the maximum information contained in the model.

3 SENSITIVITY ANALYSIS

The aim of global sensitivity analysis is to evaluate the rate of change of model output compared to variation in the model parameters[8]. Such knowledge is important for assessing the applicability of the model, determining the parameters most influential, and ultimately for understanding the behavior of the system to model. Polynomials chaos expansions once constructed offer the possibility of calculating these indices directly without the need to perform additional simulations. By exploiting the orthogonality of the basis of the sparse polynomial chaos expansion, we can calculate the mean and variance of the function as follows:

$$\overline{M} = E\left[M(\Theta)\right] = M_1$$

$$D_{pc} = Var\left[\sum_{\alpha \in A \backslash \{0\}} M_\alpha \psi_\alpha(\Theta)\right] \qquad (4)$$

Sobol indices noted $SU_{i_1...i_s}$ are defined

$$SU_{i_1,...i_s} = \sum_{\alpha \in A_{i_1...i_s}} M_\alpha^2 E\left[\psi_\alpha^2\right] / D_{pc} \qquad (5)$$

$$A_{i_1...i_s} = \left\{\alpha \in A : \alpha_k = 0 \, si \, k \notin \{i_1,...,i_s\}, \forall k = 1,...,M\right\}$$

The Sobol indices measure the contribution of an input to the output variance either solely or by interactions with the other inputs.

4 MONTE CARLO MARKOV CHAIN

A physical model can be expressed as follows:

$$M(t,\Theta) = F_{\exp}(t) + \varepsilon_t \quad t = 1,....,nt \tag{6}$$

Where $F_{\exp}$ the vector of the observed data is, M is the model describing the unsaturated-saturated flow Eqs(1-2) and $\varepsilon_t = N(0,\sigma^2)$ is the measurement error assumed independently normally distributed. Θ is the vector of parameters to be estimated from the observations. They are not directly measurable and subject to errors and uncertainty since the observations are random variables. This uncertainty is quantified in a Bayesian framework by measuring the posterior distribution which provides the relative probability of a set of parameters being plausible. We use Bayes' rule to define a posterior probability density for Θ, given an observation of the data $F_{\exp}$.

$$p\left(\Theta\middle|F_{\exp}\right) \propto p\left(F_{\exp}\middle|\Theta\right) \times p\left(\Theta\right) \tag{7}$$

where $p\left(F_{\exp}\middle|\Theta\right)$ is the likelihood function that measures how well the model fit the observed data. Under the assumption of independent, and normally distributed residual ε , the likelihood is given as:

$$p\left(F_{\exp}\middle|\Theta\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{nt}\varepsilon_i^2\right) \tag{8}$$

$p\left(\Theta\right)$ is the prior distribution of parameter, that defines the knowledge about the parameter independently of the observation. To assess parameter uncertainty of this nonlinear model and determine parameter marginal posterior distributions, Metropolis MCMC is used. It is based on the idea of generating a large number of random processes from the posterior distributions, starting from a prior distribution.

The MCMC algorithm employed in this work proceeds as follows[9]:

Initialization: Start from an initial point $\Theta^{i-1} = \Theta^0$ chosen according to the prior distribution and compute the density of probability $p\left(\Theta^0\middle|F_{\exp}\right)$ of this set of parameter.

Start iterations:

1. Generate a new point Θ^i using the multi-normal distribution as proposal distribution.

$$\Theta^i \leftarrow N\left(\Theta^{i-1},c^2 Cov\right) \tag{9}$$

Where Θ^{i-1} is the previous candidate, Cov is the covariance matrix and c is a scaling factor that depends on the dimensionality of the parameter and is used to ensure a reasonable acceptance rate. This covariance matrix is subsequently updated throughout the iterations.

$$Cov_i = \begin{cases} Cov_0 & i \le i_0 \\ Cov(\Theta^0, ..., \Theta^{i-1}) & i > i_0 \end{cases} \tag{10}$$

2. Calculate $p\left(\Theta^i \big| F_{\exp}\right)$ and compute the ratio $\alpha = \dfrac{p\left(\Theta^i \big| F_{\exp}\right)}{p\left(\Theta^{i-1} \big| F_{\exp}\right)}$

3. Randomly sample *Fact* from a uniform distribution between 0 and 1.
4. If $Fact \le \alpha$, then accept the new set of parameter, if $Fact > \alpha$, then reject the point and $\Theta^i = \Theta^{i-1}$
5. Increment i, until i is less than a prescribed number of draws *Ns*.

If the length of the chain N is sufficiently large, the obtained chain $\left\{\Theta^0, \Theta^1, ..., \Theta^N\right\}$ converges to the stationary posterior distribution of the parameter[10]. This can be very time consuming if the direct model is used. To alleviate to this problem, the polynomial chaos expansion is used instead of the direct model.

5   RESULTS AND DISCUSSIONS

5.1 Estimation of K_s and K_p using the saturated measures

In order to reduce the dimension of the model, we first estimate the parameter K_s and K_p independently from the other parameters using only observations from the saturated phase as recommended by Durner et al. (2011)[11].
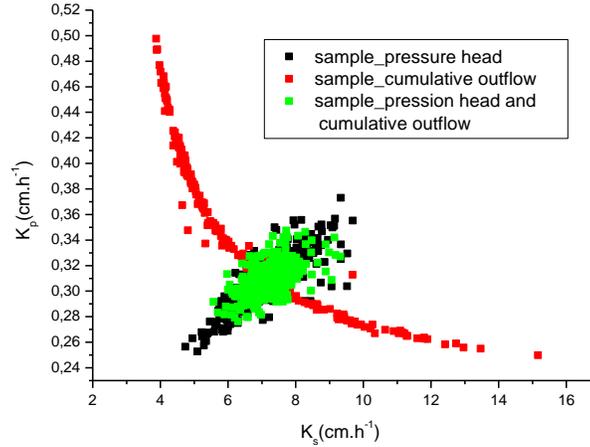

Figure 1: Scatter plot of K_s - K_p MCMC sample.

Figure 1 presents pairwise scatter plots of the posterior sample obtained with MCMC simulations. A problem of identifiability is detected in the case where only measures of

cumulative outflow are only used caused by a perfect correlation of K_s and K_p. This is justified by the fact that the cumulative outflow is actually sensitive to K_{eff} while using the pressure head, this compensation effect disappears. The 95% confidence intervals on the parameters are, $K_s = 7.087 \pm 0.76$, $K_p = 0.307 \pm 0.016$. These intervals are slightly reduced if we use both the measure of the pressure head and the cumulative outflow as shown in Figure1 .

5.2 Estimation of the Van Genuchten parameter

In this section, K_s and K_p are fixed at values obtained from estimation. A polynomial chaos expansion is then constructed according to the parameters λ, θ_r, α and n and used as a surrogate for the physical model in the inverse modeling. The global sensitivity indices are depicted in Figure 2.

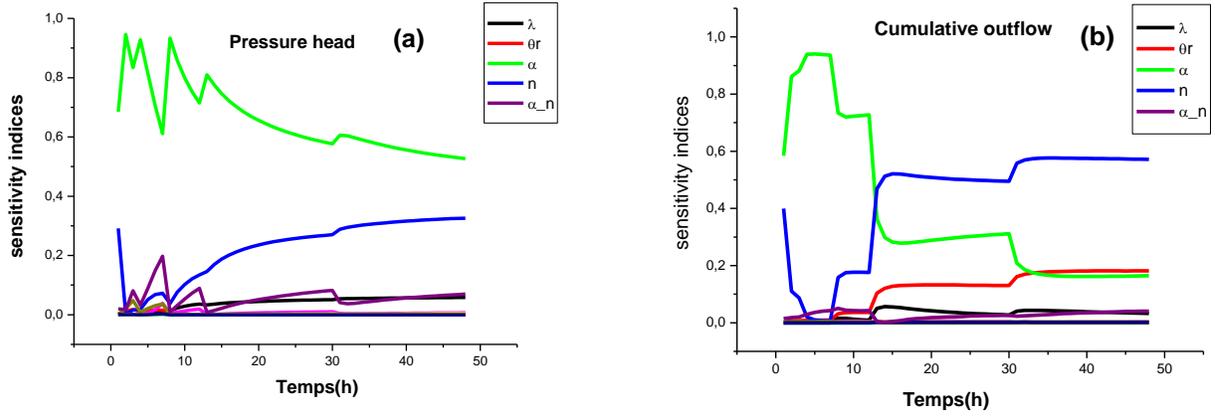Figure 2: Sensitivity indices for (a) pressure head, and (b) cumulative outflow

Figure 2a shows that the pressure head is highly sensitive to α and moderately sensitive to n. The influence of α decreases as a depression in the lower boundary is increased. Inversely, this diminution is compensated by an increase in the influence of n. Small interactions between α and n are observed. The influence of λ is small comparing to α and n. Figure 2b shows that the cumulative outflow is sensitive to both α, n. As for the pressure head, the sensitivity to α decreases after 7h whereas, conversely, the sensitivity to n and θ_r increase. Small interactions between α, n are also found. The influence of λ remains small comparing to α and n and θ_r. These results give a preliminary idea about how estimation and the uncertainty analysis is going to be. α, n could both been estimated using measures of pressure head or/and cumulative outflow. However, to estimate correctly θ_r, we must use the cumulative outflow. Using pressure head to estimate θ_r may result in a very large uncertainty.

6

Figure 3(a-b-c-d) compares marginal posterior density of parameter using only the pressure head, only cumulative outflow and both measures. Figure 3a and 3b shows that when using only the pressure head, α , n can be estimated with a good accuracy, even if the optimum values are slightly lower than their true values. Using only the cumulative outflow, the uncertainty remains reasonable but slightly wider than that obtained using only the pressure head. The optimum values are slightly higher than their true values. Using both the pressure head and the cumulative outflow does not improve the uncertainty for α , but improve it for n . Figure 3c shows that to estimate correctly λ , it is better to use both the measures of the pressure head and the cumulative outflow, as expected with sensitivity analysis. Using only one measure may cause large uncertainty. Figure 3d shows that θ_r is better estimated when using cumulative outflow but still with large uncertainty. Using the pressure head alone or combined with cumulative outflow does not give better results.
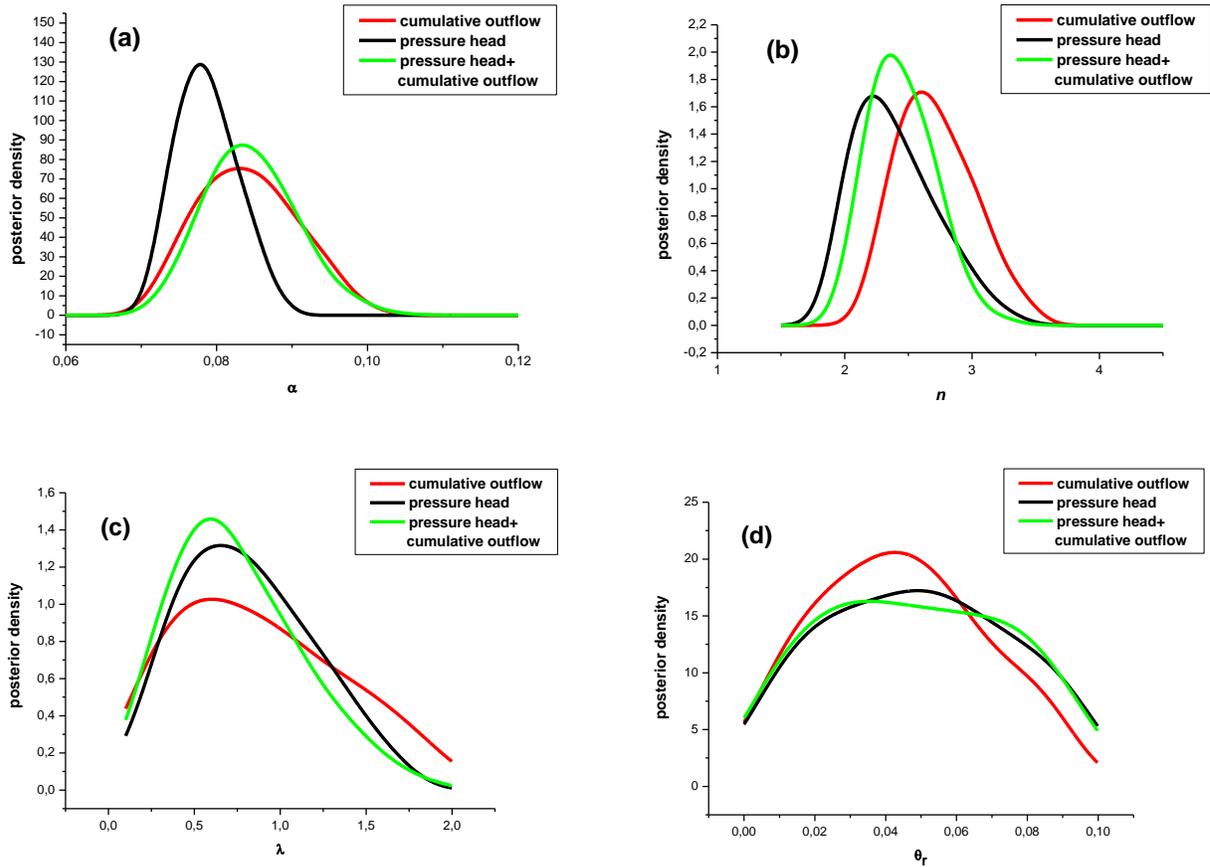

Figure 3:Posterior distribution of α , n , λ and θ_r

6   CONCLUSION

In this work, we adopt a Bayesian approach for investigating the estimation of saturated-unsaturated soil parameters synthetic case and invert the PCE in order to alleviate the computational burden. The sensitivities of pressure head and cumulative outflow to the Van Genuchten parameters are analyzed. The results show that estimation of K_s and K_p using only measures in the saturated phase is sufficiently accurate. The parameter α, n are well estimated if we use pressure head and/or cumulative outflow. To estimate θ_r, we should use the cumulative outflow. The parameter λ is not too much influential and weakly identifiable.

REFERENCES

[1] Y. Mualem,. A new model for predicting the hydraulic conductivity of unsaturated porous media.Water Resources Research 12, 513–522, (1976).

[2] M. Th. Van Genuchten and D. R. Nielsen, On describing and predicting the hydraulic properties   of unsaturated soils. Annales Geophysicae 3:615-628,(1985).

[3] C.T. Miller, C. Abhishek, M.W. Farthing,. A spatially and temporally adaptive solution of Richards' equation. Advances in Water Resources 29(4)525-45, (2006).

[4] M.D. Tocci, C.T. Kelley and C.T. Miller, Accurate and economical solution of the pressure-head form of Richards' equation by the method of lines. Advances in Water Resources 20 (1)1-14 (1997).

[5] M. Fahs, A. Younes and F. Lehmann, An easy and efficient combination of the Mixed Finite Element Method and the Method of Lines for the resolution of Richards' Equation. Environmental Modelling and Software 24 (9)1122-1126 (2009).

[6] M. A. Tatang, W. Pan, R.G. Prinn, and G.J. McRae, An efficient method for parametric uncertainty analysis of numerical geophysical models, Journal of Geophysical research 102 (. D18), 21,925-21,932, doi:10.1029/97JD01654 (1998).

[7] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering and System Safety, 93(7)964-979 (2008).

[8] T.A. Mara and Tarantola, S, Application of global sensitivity analysis of model output to building thermal simulations. Journal of Building Simulation, 1:290-302 (2008).

[9] J.A. Vrugt and W. Bouten, Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. Soil. Sci. Soc. Am. J. 66:1740-1751 (2002).

[10] A. Gelman, J.B. Carlin, H.S. Stren, and D.B. Rubin, Bayesian data analysis. Chapman and Hall, London, (1997).

[11] W. Durner and S.C. Iden, Extended multistep outflow method for the accurate determination of soil hydraulic properties near water saturation. Water Resources Research 47, W08526, (2011).