



HAL
open science

A comparison of two Bayesian approaches for uncertainty quantification

Thierry A. Mara, Frederick Delay, François Lehmann, Anis Younes

► **To cite this version:**

Thierry A. Mara, Frederick Delay, François Lehmann, Anis Younes. A comparison of two Bayesian approaches for uncertainty quantification. *Environmental Modelling and Software*, 2016, 82, pp.21-30. 10.1016/j.envsoft.2016.04.010 . hal-01310576

HAL Id: hal-01310576

<https://hal.science/hal-01310576>

Submitted on 2 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A comparison of two Bayesian approaches for uncertainty quantification

Thierry A. Mara^{a,*}, Frederick Delay^b, François Lehmann^b, Anis Younes^{b,c}

^a*PIMENT, EA 4518, Université de La Réunion, FST, 15 Avenue René Cassin, 97715
Saint-Denis, Réunion*

^b*LHyGeS, UMR-CNRS 7517, Université de Strasbourg/EOST, 1 rue Blessig, 67084
Strasbourg, France*

^c*IRD UMR LISAH, F-92761 Montpellier, France*

Abstract

Statistical calibration of model parameters conditioned on observations is performed in a Bayesian framework by evaluating the joint posterior probability density function (pdf) of the parameters. The posterior pdf is very often inferred by sampling the parameters with Markov Chain Monte Carlo (MCMC) algorithms. Recently, an alternative technique to calculate the so-called Maximal Conditional Posterior Distribution (MCPD) appeared. This technique infers the individual probability distribution of a given parameter under the condition that the other parameters of the model are optimal. Whereas the MCMC approach samples probable draws of the parameters, the MCPD samples the most probable draws when one of the parameters is set at various prescribed values. In this study, the results of a user-friendly MCMC sampler called DREAM_(ZS) and those of the MCPD sampler are compared. The differences between the two approaches are highlighted before running a comparison inferring two analytical distributions with collinearity and multimodality. Then, the performances of both samplers are compared on an artificial multistep outflow experiment from which the soil hydraulic

*Corresponding author: mara@univ-reunion.fr

parameters are inferred. The results show that parameter and predictive uncertainties can be accurately assessed with both the MCMC and MCPD approaches.

Keywords: Bayesian parameter estimation, parameter uncertainty, predictive uncertainty, MCPD sampler, DREAM_(ZS) MCMC, soil hydraulic parameter identification

Contents

1	Introduction	4
2	Inverse Problem	6
2.1	Bayesian inference	6
2.2	Maximal conditional posterior distribution	7
3	Parameter uncertainty assessment	8
3.1	The DREAM _(ZS) MCMC sampler	8
3.2	The MCPD sampler	10
4	Computational issues	13
4.1	MCPD versus MCMC draws	13
4.2	Predictive uncertainty	14
5	Numerical exercises	15
5.1	A 10-dimensional twisted Gaussian target distribution	15
5.2	An 11-dimensional multimodal target density	18
5.3	Identification of soil hydraulic parameters	20
5.3.1	Synthetic drainage experiment	20
5.3.2	Results and discussion	23
6	Conclusions	24

1. Introduction

The validation of computer models is an essential task to increase their credibility. One of the most important exercises in the validation framework is to check whether the computer model adequately represents reality (Bayarri et al., 2007). This is achieved by comparing model predictions to observation data. This exercise generally leads to model calibration because the model parameters are usually poorly known a priori (*i.e.* before collecting data). Good practice in calibration of computer models consists of searching for all parameter values that satisfactorily fit the data, thus determining their plausible range of uncertainty. This can be achieved in a Bayesian framework in which the prior knowledge about the model and the observed data are merged to define the joint posterior probability distribution function (pdf) of the parameters. The issue is then to assess the joint posterior pdf.

The inference of model parameter posterior pdf by means of Markov chain Monte Carlo (MCMC) sampling techniques (Metropolis et al., 1953; Hastings, 1970) has received much attention in the last two decades. MCMC explores the region of plausible values in the parameter space and provides successive parameter draws directly sampled from the target joint pdf. Some selection criteria are used to ensure that the successive draws in the chain improve. This means that, throughout the sampling process, probable draws with respect to the target distribution are more likely drawn. Many developments and improvements have been proposed to accelerate MCMC convergence.

Grenander and Miller (1994) developed the Langevin MCMC, which accelerates the convergence of the chains by exploiting the Jacobian of the target distribution. This MCMC sampler may require that the computer model

provide the local sensitivities to compute the Jacobian of the target distribution. In practice, modelers generally estimate the gradient by finite differences via a surrogate or coarse-scale model to alleviate the computational burden (see for instance, [Dostert et al., 2009](#); [Angelikopoulos et al., 2015](#)). [Haario et al. \(2006\)](#) developed the Delayed Rejection Adaptive Metropolis (DRAM), an algorithm that increases the rate of acceptance of MCMC draws by exploiting the delayed rejection trick proposed in [Tierney and Mira \(1999\)](#) and the adaptive Metropolis algorithm of [Haario et al. \(2001\)](#). [ter Braak and Vrugt \(2008\)](#) developed the Differential Evolution-Markov Chain (DE-MC) algorithm, which merges the differential evolution method of [ter Braak \(2006\)](#) and the Shuffled Complex Evolution Metropolis (SCEM) method proposed by [Vrugt et al. \(2003\)](#). DREAM improves the efficiency of MCMC by running multiple chains in parallel for a wider and quicker exploration of the parameter space in addition to a self-adaptive randomized subspace sampling ([Vrugt et al., 2009](#)). Recently, the algorithm of DREAM has been embedded in UCODE_2014, dedicated to inverse modeling ([Lu et al., 2014](#)). [Laloy and Vrugt \(2012\)](#) then developed $\text{DREAM}_{(ZS)}$, that ensures convergence with fewer chains in parallel than DREAM.

Recently, [Mara et al. \(2015\)](#) proposed a new probabilistic approach to the inverse problem whose main idea is to maximize the joint posterior pdf of a parameter set with one selected parameter sampling successive prescribed values. This provides the so-called Maximal Conditional Posterior Distribution (MCPD) of the selected parameter. The main advantage of the recent MCPD technique is that parameter distributions can be inferred independently. Therefore, the MCPDs can be simultaneously evaluated on multicore computers (or on multiple computers). This drastically reduces the computational effort in terms of computational time units (CTU).

The MCPD and MCMC samplers assess the same target distribution, namely, the parameter joint posterior pdf. Nevertheless, the two samplers do not provide the same results. In general, the MCPD of a single parameter does not correspond to its marginal posterior distribution. In addition, the MCPD sampler only provides a few sets of probable draws while MCMC generates a large number of draws sampled in agreement with the target distribution. Nevertheless and as advocated in this study, both samplers are valuable Bayesian methods for statistical inverse problems. Hence, the main objective of the present work is to compare the ability of MCPD and DREAM_(ZS) MCMC samplers to quantify model output and model parameter uncertainties.

The paper is organized as follows: Section 2 summarizes the inversion in a Bayesian framework and recalls the principles of the recent MCPD technique. The general algorithms ruling the DREAM_(ZS) MCMC and MCPD samplers are introduced in Section 3. In Section 4, we discuss on the analogy and the differences between MCPD and MCMC draws. Section 5, emphasizes the comparison between MCMC and MCPD samplings: 1) for the inversion of multimodal and correlated functions, and 2) for the evaluation of soil hydraulic properties from a synthetic one-dimensional drainage experiment. Finally, a summary with conclusions is presented in Section 6.

2. Inverse Problem

2.1. Bayesian inference

In probabilistic inverse modeling, the parameter set $\mathbf{x} = (x_1, \dots, x_d)$ of a computer model is inferred from a set of observation data \mathbf{y} using the Bayesian inference, which defines the conditional joint posterior pdf as fol-

lows:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \quad (1)$$

where $p(\mathbf{x})$ is the prior density that characterizes the investigator's beliefs about the parameters before collecting the new observations, and $p(\mathbf{y}|\mathbf{x})$ is the likelihood function, which measures how well the model fits the data.

The parameter set that maximizes Eq. (1), namely:

$$\mathbf{x}^{\text{MAP}} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y}), \quad (2)$$

is called the Maximum A Posteriori (MAP) estimate of the parameters. It is the most probable parameter set given the data and can be inferred via an optimization technique. The marginal posterior pdf that characterizes the uncertainty of a single parameter is defined by the following integral:

$$p(x_i|\mathbf{y}) = \int p(\mathbf{x}|\mathbf{y})d\mathbf{x}_{-i}, \quad \forall i = 1, \dots, d \quad (3)$$

where \mathbf{x}_{-i} represents all the parameters except x_i . Usually, the integral in Eq. (3) is evaluated by a multidimensional quadrature method or by direct summations in a large sample of $p(x_i|\mathbf{y})$ obtained, for instance, via an MCMC technique.

2.2. Maximal conditional posterior distribution

[Mara et al. \(2015\)](#) define the maximal conditional posterior distribution of x_i as follows:

$$\mathcal{P}(x_i) = \max_{\mathbf{x}_{-i}} (p(\mathbf{x}_{-i}|\mathbf{y}, x_i)) \times p(x_i|\mathbf{y}). \quad (4)$$

An informal definition can be given by stating that a point estimate of the MCPD is the maximal value reached by the joint pdf Eq. (1) for a given (prescribed) value of one parameter (*i.e.* x_i). This maximal value, in the context

of model inversion, assumes that the set \mathbf{x}_{-i} maximizes Eq. (1), knowing that x_i is prescribed. By applying the axiom of conditional probabilities to Eq. (4), it can be stated that $\max \{p(\mathbf{x}_{-i}|\mathbf{y}, x_i)\} \times p(x_i|\mathbf{y}) = \max_{\mathbf{x}_{-i}} \{p(\mathbf{x}_{-i}, x_i|\mathbf{y})\}$. Therefore, the MAP estimate (when it exists) belongs to the MCPD of all parameters.

In view of the MCPD definition, especially its interpretation in terms of the x_i draws for the other parameters at their optimal values, the MCPD can provide information on the uncertainty attached to a single parameter. Obtaining uncertainties for all parameters is simply achieved by calculating the individual MCPD of all parameters.

3. Parameter uncertainty assessment

3.1. The $DREAM_{(ZS)}$ MCMC sampler

The MCMC samplers generate successive draws of parameter sets that converge toward the posterior density $p(\mathbf{x}|\mathbf{y})$. Several methods are reported in the literature (*e.g.* Grenander and Miller, 1994; Haario et al., 2006; Vrugt et al., 2009; Laloy and Vrugt, 2012), but they all rely on the Metropolis-Hasting algorithm, which proceeds according to the following schedule:

- (i) Choose an initial estimate of the parameter set \mathbf{x}^0 and a proposal distribution $q(\mathbf{a}, \mathbf{b})$ that randomly derives the parameter set \mathbf{a} from an input \mathbf{b} .
- (ii) From the current set \mathbf{x}^k , generate a new candidate \mathbf{x}^* with the generator $q(\mathbf{x}^*, \mathbf{x}^k)$.
- (iii) Compute $\alpha = p(\mathbf{x}^*|\mathbf{y})p(\mathbf{x}^* \leftarrow \mathbf{x}^k)/p(\mathbf{x}^k|\mathbf{y})p(\mathbf{x}^k \leftarrow \mathbf{x}^*)$, where $p(\mathbf{b} \leftarrow \mathbf{a})$ is the transition probability from individual \mathbf{a} to individual \mathbf{b} associated

with the generator q . Additionally, draw a random number $u \in [0, 1]$ from a uniform distribution.

- (iv) If $\alpha \geq u$, set $\mathbf{x}^{k+1} = \mathbf{x}^*$, otherwise, set $\mathbf{x}^{k+1} = \mathbf{x}^k$.
- (v) Resume from (ii) until the chain $\{\mathbf{x}^0, \dots, \mathbf{x}^k\}$ converges or a prescribed number of iterations k_{max} is reached.

The calculation of $p(\mathbf{x}^*|\mathbf{y})$ in (iii) requires that the forward model be run for the set of parameters \mathbf{x}^* . This is the most expensive computational step. The crux step at the origin of the computation costs is the step (iv), which may reject many candidates. The choice of the generator $q(\cdot)$ is a key feature for the acceptance rate.

If the Markov chain is constructed correctly, the chain $\{\mathbf{x}^0, \dots, \mathbf{x}^k\}$ should converge for its last elements toward the targeted posterior distribution $p(\mathbf{x}|\mathbf{y})$ (Geyer, 1992; Robert and Casella, 2004). Unfortunately, the number of necessary draws cannot be guessed in advance. In practice, one regularly evaluates the \hat{R} -statistic of Gelman and Rubin (1992) and decides whether to stop the sampling procedure. Usually, the first draws of the chains are overlooked because they correspond to a so-called burn-in period within which the Markov process wanders in the entire parameter space. Most of the improvements brought to the MCMC samplers in the last decade aimed at diminishing this burn-in period.

In this study, we use the DREAM_(ZS) software developed by Laloy and Vrugt (2012). The algorithm relies on a multiple-chain method that computes different trajectories (sub-chains) in parallel to better explore the parameter space when seeking the target posterior pdf. A new candidate for each chain is drawn from an archive of past states (denoted \mathbf{Z}) by using the differential evolution algorithm and a snooker updater developed by ter Braak and Vrugt

(2008). These two algorithms ensure that the new candidates have the desired scale and orientation. Note that the sample of past states \mathbf{Z} plays the role of the generator q (see above) and is periodically archived and updated. $\text{DREAM}_{(ZS)}$ also uses a self-adaptive randomized subspace sampling and explicitly discards aberrant trajectories in the parameter space. The initial candidates can be sampled from any desired distributions. In particular, the user can impose an initial proposal distribution q like we did in the first two numerical exercises in § 5. The interested readers are referred to [Vrugt \(2016\)](#) for more details about the use and implementation of DREAM and $\text{DREAM}_{(ZS)}$.

3.2. The MCPD sampler

As compared to the MCMC sampler, the MCPD sampler is thoroughly described because it is a very recent approach. The algorithm used to compute the MCPDs is divided into three parts: in part 1 (step (i) below), all the probable optima of $p(\mathbf{x}|\mathbf{y})$ are investigated. In part 2 (steps (ii-v) below), the MCPD of the current parameter around each probable optimum is roughly estimated. In part 3, the discretization of the MCPD is refined. It is worth noting that in parts 2 & 3, because the conditional optimizations are performed around a local optimum, it is assumed that there is only a single optimum. The algorithm proceeds as follows:

- (i) Find by optimization all probable optimal candidates (including the MAP estimate) $\{\mathbf{x}^{\text{opt},1}, \dots, \mathbf{x}^{\text{opt},M}\}$ — that is, all local optima of $p(\mathbf{x}|\mathbf{y})$ according to the following criterion: $p(\mathbf{x}^{\text{opt},m}|\mathbf{y})/p(\mathbf{x}^{\text{MAP}}|\mathbf{y}) > 0.01$. with $\mathbf{x}^{\text{MAP}} = \arg\max\{p(\mathbf{x}^{\text{opt},m}|\mathbf{y}), m = 1, \dots, M\}$. Set $i = 1$ for the current parameter x_i and define a maximum number of iterations for the refinement of the MCPDs, *e.g.* $N_{\text{it}} = 10$.

- (ii) Set $m = 1$ for the current optimal candidate $\mathbf{x}^{\text{opt},m}$,
- (iii) Set the iteration number $k = 1$ and the initial discretization step $\Delta = 0.10$.
- (iv) Set $x_i^\pm = x_i^{\text{opt},m}(1 \pm k\Delta)$ and find $\mathbf{x}^\pm = (\mathbf{x}_{-i}^\pm, x_i^\pm)$ by solving Eq. (4). This is achieved by maximizing $p(\mathbf{x}_{-i}|\mathbf{y}, x_i^*)$ for $x_i^* = x_i^+$ and $x_i^* = x_i^-$ successively.
- (v) If $p(\mathbf{x}^\pm|\mathbf{y})/p(\mathbf{x}^{\text{MAP}}|\mathbf{y}) > 0.01$, set $k = k + 1$ and go to iv); else, set $it = 1$ and continue
- (vi) Denote $\{\mathbf{x}^{k_i}, \mathcal{P}(x_i^{k_i}) = p(\mathbf{x}^{k_i}|\mathbf{y}), k_i = 1, \dots, k\}$; the first draws sorted as $x_i^{k_i} < x_i^{k_i+1}$. Refine the current MCPD sampling by finding $\mathbf{x}^{k+1} = (\mathbf{x}_{-i}^{k+1}, x_i^{k+1})$ such that $\mathbf{x}_{-i}^{k+1} = \underset{\mathbf{x}_{-i}}{\text{argmax}} p(\mathbf{x}_{-i}|\mathbf{y}, x_i^{k+1})$, with

$$\begin{cases} x_i^{k+1} &= \frac{x_i^{k_m+1} + x_i^{k_m}}{2} \\ k_m &= \underset{k_i}{\text{argmax}} \{|\mathcal{P}(x_i^{k_i+1}) - \mathcal{P}(x_i^{k_i})|\} \end{cases} \quad (5)$$
- (vii) Set $k = k + 1$ and add $(\mathbf{x}^{k+1}, \mathcal{P}(x_i^{k+1}))$ to the subset $\{\mathbf{x}^{k_i}, \mathcal{P}(x_i^{k_i}), k_i = 1, \dots, k\}$. If $it < N_{\text{it}}$, set $it = it + 1$ and go to vi); otherwise, continue
- (viii) Set $m = m + 1$ (change of local optimum). If $m \leq M$ go to iii); otherwise continue
- (ix) Set $i = i + 1$ (next parameter) and if $i \leq d$ resume from (ii); otherwise, stop.

Unlike MCMC, the MCPD relies on an optimization technique in steps (i), (iv) and (vi). The computational effort of the algorithm resides in these steps. In the numerical exercises below, the optimizations are performed

with a gradient-based algorithm. Nevertheless, MCPD calculations are compatible with any optimization algorithm and could use, for instance, global optimization techniques that do not rely on gradients. A gradient-based method requires that an initial guess of the solution \mathbf{x}^0 be provided beforehand. To find all local optima, we systematically perform $N_{\text{init}} = 20$ optimization procedures with different initial solutions randomly drawn within the parameters' prior (*i.e.* $\mathbf{x}^0 \sim p(\mathbf{x})$).

It is worth noting that the convergence of gradient-based algorithms accelerates if the Jacobian of the target pdf is also provided. The number of initial optimizations N_{init} is a matter of choice as well as the number of refinement points N_{it} . They condition the total computational costs. Our experience suggests that $N_{\text{it}} = 10$ is sufficient to obtain accurate results (see also [Mara et al., 2015](#)).

The rough estimation of the current MCPD is performed in the second part of the algorithm (from (ii) to (v)). In step (v), at the first iteration ($k = 1$) for the current variable x_i , if $p(\mathbf{x}^\pm | \mathbf{y}) \approx p(\mathbf{x}^{\text{MAP}} | \mathbf{y})$, it can mean that $\mathcal{P}(x_i)$ is flat. It is then recommended that the size step be increased (*e.g.* set $\Delta = 1.5 \times \Delta$). Conversely, if $p(\mathbf{x}^\pm | \mathbf{y}) \approx 0$, $\mathcal{P}(x_i)$ can be very narrow and one must decrease the current size step (*e.g.* set $\Delta = \Delta/1.5$).

The strategy to refine the MCPD assessment given in Eq. (5) is illustrated in Figure 1. After the second part of the algorithm, one obtains a rough discretization of the MCPD (in circles). Then, according to Eq. (5), the next prescribed value of x_i (*i.e.* x_i^{k+1}) is determined. The latter is chosen where the gap between two successive MCPD values is maximal. In addition to the above schedule, one can mention that the algorithm is set up to sample the values of the parameter x_i in the vicinity of each probable local optimum. In complex problems, it is not ensured that any optimization technique will

retrieve all local optima. However, some previous tests showed that when a local optimum has been missed, it is often retrieved in steps (iv) or (vi) when sampling the MCPDs by optimizing the conditional pdf $p(\mathbf{x}_{-i}|\mathbf{y}, x_i)$.

Insert Figure 1 about here

4. Computational issues

4.1. MCPD versus MCMC draws

By definition, the MCMC and MCPD samplers provide different results. Nevertheless, both approaches assess the same target distribution, namely, the parameter posterior pdf $p(\mathbf{x}|\mathbf{y})$. The MCMC sampler provides a large set of candidates and their associated weights $(\mathbf{x}_{\text{MCMC}}, p(\mathbf{x}_{\text{MCMC}}|\mathbf{y}))$ while the MCPD sampler only provides a small set of draws $(\mathbf{x}_{\text{MCPD}}, p(\mathbf{x}_{\text{MCPD}}|\mathbf{y}))$. The MCMC draws \mathbf{x}_{MCMC} represent a stochastic sample of the parameter values distributed with respect to $p(\mathbf{x}|\mathbf{y})$. The MCPD sample \mathbf{x}_{MCPD} is a set of probabilistic draws of the parameter values.

Regarding the implementation of the two algorithms, the MCMC DREAM_(ZS) sampler is much easier to plug into a given computer model. DREAM_(ZS) does not need for modifications of the computer model but only requires that the target distribution be defined. The efficiency of the MCPD sampler is enhanced if the partial derivatives (of the model response w.r.t. the parameters) are also provided by the computer model. Otherwise, the use of the finite-differences approach to estimate the partial derivatives deteriorates the performance of the MCPD calculations. Typically, the number of model calls is multiplied by the number of parameters. This is an important feature to be aware of before using the MCPD approach for inverse problems. In the following numerical exercises, the partial derivatives are systematically com-

puted by the forward models. This slightly increases the computational time of one model run.

The MCPD of a given parameter, say $\mathcal{P}(x_i)$, does not always match its posterior marginal pdf $p(x_i|\mathbf{y})$. This only happens when $\max(p(\mathbf{x}_{-i}|\mathbf{y}, x_i))$ is constant for all prescribed values of x_i around the current optimum. In that case, it can be shown that $\mathcal{P}(x_i) \propto p(x_i|\mathbf{y})$ (see [Mara et al., 2015](#)). It is worth specifying that this invariance of $\max(p(\mathbf{x}_{-i}|\mathbf{y}, x_i))$ with respect to the value of x_i does not mean that $p(\mathbf{x}_{-i}|\mathbf{y}, x_i)$ is independent of x_i . Let us consider the example of the following target density: $p(x_1, x_2|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, with \mathcal{N} the bi-Gaussian density, $\boldsymbol{\mu}_0$ the vector of means and $\boldsymbol{\Sigma}$ a given non-diagonal covariance matrix. It is obvious that $p(x_2|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_1)$ depends on the value of x_1 . However, $\max(p(x_2|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_1)) = 1/2\pi\sqrt{\det \boldsymbol{\Sigma}}$ for any value of x_1 . Thus, for the considered target density, the posterior pdf of x_i , $i = 1, 2$, matches its maximal conditional posterior density defined as,

$$\pi(x_i) = \frac{\mathcal{P}(x_i)}{\int_{-\infty}^{+\infty} \mathcal{P}(x_i) dx_i}. \quad (6)$$

From the MCPD draws, the integral in Eq. (6) is computed with the Simpson quadrature rule. By defining the vector of normalized weights as, $\mathbf{w}^* = p(\mathbf{x}^*|\mathbf{y}) / \int_{-\infty}^{+\infty} \mathcal{P}(x_i) dx_i$, the maximal conditional posterior densities ($\mathbf{x}_{\text{MCPD}}, \mathbf{w}_{\text{MCPD}}$) and the MCMC draws ($\mathbf{x}_{\text{MCMC}}, \mathbf{w}_{\text{MCMC}}$) can be plotted on the same graph.

4.2. Predictive uncertainty

To obtain the predictive posterior density of an observation data y^* given the dataset at hand \mathbf{y} , the following integral must be calculated:

$$p(y^*|\mathbf{y}) = \int p(y^*|\mathbf{y}, \mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x} \quad (7)$$

$p(y^*|\mathbf{y}, \mathbf{x})$ measures how likely the model response value is y^* given the set of parameters \mathbf{x} and the dataset \mathbf{y} . This integral merges the likelihood function

and the parameter joint posterior pdf. Evaluating Eq. (7) allows assigning uncertainty bounds to the model predictions. Assume that the assessment of the MCPDs provides n draws, then with the MCPD approach Eq. (7) is approximated by:

$$\hat{p}(y^*|\mathbf{y}) = \frac{\sum_{k=1}^n w_{\text{MCPD}}^k p(y^*|\mathbf{y}, \mathbf{x}_{\text{MCPD}}^k)}{\sum_{k=1}^n w_{\text{MCPD}}^k}. \quad (8)$$

with $\mathbf{x}_{\text{MCPD}}^k$ the k^{th} MCPD draw.

5. Numerical exercises

5.1. A 10-dimensional twisted Gaussian target distribution

For this first numerical exercise, we target the twisted Gaussian distribution proposed by [Haario et al. \(1999\)](#) given by,

$$p(\mathbf{x}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) \propto \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}),$$

where $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ is the ten-dimensional Gaussian distribution (*i.e.* $\mathbf{x} = (x_1, \dots, x_{10})$) with mean $\boldsymbol{\mu}_0 = (0, -0.1x_1^2 + 10, 0, \dots, 0)$ and covariance $\boldsymbol{\Sigma} = \text{diag}(100, 1, \dots, 1)$.

This target distribution is very challenging for both MCPD and MCMC samplers because of the nonlinear (banana-shaped) relationship between (x_1, x_2) . The target distribution has only one optimum; that is, $\mathbf{x}^{\text{MAP}} = (0, 10, 0, \dots, 0)$. Note that when $\mathcal{P}(x_2)$ is assessed, the maximization of the conditional distribution $p(\mathbf{x}_{-2}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_2)$ can return three local optima: $x_1 \in \{-\sqrt{100 - 10x_2 - 0.5}, 0, \sqrt{100 - 10x_2 - 0.5}\}$. Part 2 of the MCPD algorithm described above allows retrieval of only one of them (depending on the initial guess) because it is assumed, in this part, that the conditional distributions have only one optimum. Hence, the MCPD sampler may fail at inferring $\mathcal{P}(x_2)$ directly. However, evaluating the MCPD of x_1 —that is, maximizing $p(\mathbf{x}_{-1}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_1)$ —

gives $x_2 = 10 - 0.1x_1^2$. Thus, one obtains the banana-shaped relationship between the two variables and can infer $\mathcal{P}(x_2)$ subsequently.

The results are depicted in Figure 2. They confirm that the MCPD sampler fails to infer the true MCPD of x_2 (broken red line) because the conditional maximization around the MAP estimate provides $x_1 = 0$, which is a local optimum when $x_2 < 10$. However, as expected, $\mathcal{P}(x_2)$ can be inferred from the conditional maximization of $p(\mathbf{x}_{-1}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_1)$ (continuous red line). This result supports the idea that when an optimum is missed, it can be retrieved during the other optimization steps.

The pairwise analysis of the MCPD draws reveals the banana-shaped relationship between x_1 and x_2 (row #2 column #1 in Figure 2). Two curves are depicted for (x_1, x_2) , the first corresponding to the optimal sought values of the parameter x_2 for the sampled (prescribed) values of x_1 , and the second one corresponding to optimal values of x_1 for sampled x_2 . The MCPD of the other parameters (x_3, \dots, x_{10}) are found approximately Gaussian (MCPDs beyond x_3 are not reported in Figure 2). An analysis of their pairwise scatterplots does not reveal other correlation structures. In fact, like the scatterplot of (x_1, x_3) , one observes two orthogonal lines that represent the optimal values of x_i versus prescribed values of x_j and optimal x_j versus prescribed x_i , respectively. They indicate that prescribing x_i and maximizing $p(\mathbf{x}_{-i}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_i)$ always provides $\mathbf{x}_{-i}^{\text{MAP}}$ (and vice versa).

Insert Figure 2 about here

This numerical exercise with the MCPD algorithm required approximately 900 model calls to find the MAP estimate. We recall that, for this purpose, the optimization program was repeated $N_{\text{init}} = 20$ times with different initial guesses. The conditional optimizations for inferring the MCPDs required around 1,000 extra model calls. Distributing the independent searches

of the MAP and then the independent calculations of the ten distinct MCPDs over ten computers (or ten CPU cores) would yield an inexpensive total number of computational time units (CTU) of approximately 190.

We also assessed $p(\mathbf{x}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ with $\text{DREAM}_{(ZS)}$. For a fair comparison with the MCPD approach, we launched $N_c = 10$ chains simultaneously. With $\text{DREAM}_{(ZS)}$, it is possible to impose an initial proposal distribution. The choice of the proposal distribution can have a substantial impact on the length of the burn-in period. Following the work of [Vrugt et al. \(2009\)](#), we chose: $q(\mathbf{a}, \mathbf{b}) = \mathcal{N}(\mathbf{b}, 5\mathbf{I}_{10})$, with \mathbf{I}_{10} representing the 10-dimensional identity matrix. First, a total number of 10,000 CTUs was chosen, which corresponds to a total of $10 \times 10,000$ model calls. As reported in [Figure 3](#), the \hat{R} -statistic of x_1 , x_2 and x_3 show that the overall chains have converged after 20,000 runs.

[Figure 2](#) (diagonal plots) shows that the MCMC draws are located, as expected, below the MCPD envelope. MCMC samples probable solutions, when MCPD only seeks solutions that maximize the target distribution conditioned onto one of the parameters (and the data). We note that the MCMC draws of each parameter are spread over the uncertainty range delimited by the MCPD draws. Both samples seem to satisfactorily represent the salient feature of the target distribution. The comparison of the estimated densities also indicates a good agreement between the two approaches, except for x_2 (see [Figure 2](#) row #2, column #2). As already mentioned, $\pi(x_2)$ matches the marginal posterior pdf $p(x_2|\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ if $\max(p(\mathbf{x}_{-2}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_2))$ is constant. However, it can be proven that

$$\max(p(\mathbf{x}_{-2}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, x_2)) \propto e^{-\frac{1}{2}(-x_2^2 - 0.1x_2)},$$

which depends on the value of x_2 . Consequently, the MCPD of x_2 does not match its marginal pdf. This explains the difference between the two

densities in Figure 2 (row #1, column #2).

Insert Figure 3 about here

5.2. An 11-dimensional multimodal target density

In this exercise, we consider the following multimodal target density:

$$p(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \mathbf{C}) = \frac{1}{6}\mathcal{N}(\boldsymbol{\mu}_1, 5\mathbf{C}) + \frac{2}{6}\mathcal{N}(\boldsymbol{\mu}_2, 5\mathbf{I}_d) + \frac{3}{6}\mathcal{N}(\boldsymbol{\mu}_3, 5\mathbf{I}_d) \quad (9)$$

where $\mathcal{N}(\boldsymbol{\mu}_i, 5\mathbf{I}_d)$ is the multi-Gaussian density of mean vector $\boldsymbol{\mu}_i$ and covariance $5\mathbf{I}_d$. \mathbf{I}_d , the d -dimensional identity matrix, indicates that the parameters (x_1, \dots, x_d) are independent in the second and third Gaussian densities in Eq. (9).

\mathbf{C} is a correlation matrix with null off-diagonal elements except for $C_{1,2} = C_{2,1} = -0.5$ and $C_{1,3} = C_{3,1} = 0.8$. These non-null terms impose, for the first Gaussian density in Eq. (9), a negative correlation between x_1 and x_2 and a strong positive correlation between x_1 and x_3 . The three modes of each parameter are grouped in the vectors of means $\boldsymbol{\mu}_1 = (-5, \dots, 5)$, $\boldsymbol{\mu}_2 = (1, \dots, 11)$ and $\boldsymbol{\mu}_3 = (11, \dots, 1)$.

In Mara et al. (2015), the MCPD sampler was faced with a similar target density with $d = 25$. It was shown that the MCPD did not match the marginal pdf because the different Gaussian densities in Eq. (9) overlapped. Here, we consider a mildly dimensional case by setting $d = 11$. With DREAM_(ZS), eleven chains in parallel were run simultaneously to infer the target density in a maximal prescribed number of 10,000 CTUs.

Figure 4 reports on the draws from the MCPD and MCMC samplers for parameters x_1 , x_2 and x_3 . Both samplers were able to retrieve the three modes. As noted in the previous exercise, the MCMC draws are spread beneath the MCPD envelope (diagonal plots of Figure 4). Figure 5 shows

that the three modes were located after about 3,000 runs, a relatively long burn-in period. With the eleven chains, $\text{DREAM}_{(ZS)}$ identified two modes of three at the beginning of the search (for x_1 these two modes are $(-5, 11)$ see Figure 5), then it took about 3,000 CTUs to find the last mode ($x_1 = 1$). This burn-in period could have been reduced by increasing the number of chains or by imposing the three modes as initial candidates.

The correlation structure between (x_1, x_2) and (x_1, x_3) for the mode associated with $\boldsymbol{\mu}_1$ is confirmed by the MCPD draws. The off-diagonal plot in row #2 and column #1 of Figure 4 shows that the MCPD draws close to $\boldsymbol{\mu}_1$ are located upon two non-orthogonal lines. We remind that these lines are the optimal values of x_i for the prescribed values of x_j and optimal x_j for the prescribed x_i . The negative slopes for the pair (x_1, x_2) indicates the negative correlation between the parameters. Conversely, the pair (x_1, x_3) shows positive correlation (row #3, column #1 in Figure 4). Despite the fact that MCMC and MCPD provide different results (here, MCPD does not match the marginal pdf), the MCPD sampler is able to assess the posterior uncertainty range of the parameters which is an important feature of model inversion.

Finally, it is worth mentioning that the MCPD sampler took about 140 runs to find the three modes and about 1,600 additional model calls to evaluate all MCPDs in a sequential calculation. With 11 parallel sessions (for 11 parameters), the CTU would have been approximately 156 runs, which is few compared with the 3,000 CTU required by the burn-in period of MCMC (see above). This result is conducive to perform a preliminarily search for all local optima (as for the MCPD) before running $\text{DREAM}_{(ZS)}$. This should alleviate the computational burden of the MCMC sampler by reducing the burn-in period.

Insert Figure 4 about here

Insert Figure 5 about here

5.3. Identification of soil hydraulic parameters

5.3.1. Synthetic drainage experiment

Characterizing the hydraulic properties of soils is crucial to predict groundwater resources in aquifers and forecast the future of contaminants in the soil. Multistep outflow drainage experiments are usually conducted to estimate these parameters (van Dam et al., 1994; Eching et al., 1994; Vrugt and Bouten, 2002; Durner and Iden, 2011). In these experiments, a flow cell filled with a saturated soil is drained by imposing multistep negative pressure heads at the lower boundary of the column. The experimental device is generally equipped with a tensiometer that measures the pressure head during the experiment. The outflow volume of water is monitored automatically with an electronic balance. Inverse modeling consists in identifying the soil hydraulic properties from these measurements.

The flow through the porous medium is governed by the nonlinear one-dimensional Richard's equation:

$$\frac{\partial \omega}{\partial t} = \frac{\partial}{\partial z} \left[K(h) \left(\frac{\partial h}{\partial z} - 1 \right) \right], \quad (10)$$

where t (min) is time, z (cm) is the vertical coordinate (positive downward), and K (cm.min⁻¹) is the unsaturated hydraulic conductivity. The water content ω (cm³.cm⁻³) and the pressure head h cm are the state variables. In the present work, $K(h)$ is modeled by the Mualem-van Genuchten (MvG) retention curve (Mualem, 1976; van Genuchten, 1980),

$$K(S_e) = k_s \times S_e^\lambda \left(1 - (1 - S_e^{1/m})^m \right)^2, \quad (11)$$

where k_s (cm.min⁻¹) is the saturated hydraulic conductivity, and S_e (-) is the effective saturation defined as follows:

$$S_e = \frac{\omega - \omega_r}{\omega_s - \omega_r} = \begin{cases} \frac{1}{(1 + |\alpha h|^n)^m} & h < 0 \\ 1 & h \geq 0 \end{cases} \quad (12)$$

where $m = 1 - 1/n$. The soil hydraulic parameters are the saturated hydraulic conductivity k_s (cm.min⁻¹), the saturated water content ω_s (cm³.cm⁻³), the residual water content ω_r (cm³.cm⁻³) and the MvG fitting coefficients α (cm⁻¹), n (-) and λ (-).

Eqs. (10-12) are solved with a standard Galerkin finite element method in conjunction with the Newton linearization method associated with the primary variable switching method (Diersch and Perrochet, 1999; Hayek et al., 2008). An implicit time scheme is used. The calculation of the partial derivatives matrix $\partial h / \partial \mathbf{x}$ is computed analytically by solving the sensitivity equations of the discretized direct problem. The program also computes the partial derivatives of the average water content with respect to the unknown parameters at each time step (*i.e.* $\partial \bar{\omega} / \partial \mathbf{x}$). The latter allows for the fast convergence of the optimization procedure used in the MCPDs assessment.

We model a laboratory multistep outflow drainage experiment of a column of length $L = 6$ cm and diameter $D = 8.5$ cm. Synthetic data are obtained by running the flow model for a given input parameter set and noising the model responses with independent Gaussian random noises. The responses of interest that are used in the inverse modeling are the pressure head h at 3 cm below the top of the column (the corresponding noisy data is denoted \mathbf{y}_h) and the average soil water content $\bar{\omega}$ (data denoted $\mathbf{y}_{\bar{\omega}}$) obtained from the cumulative outflow using the initial water content. The data are depicted in Figure 6.

In the present application, the two data series ($\mathbf{y}_h, \mathbf{y}_{\bar{\omega}}$) have been cor-

rupted (see above) for each observation with an additive independent and normally distributed error with variances $(\sigma_h^2)^{ex} = 4 \text{ (cm}^2\text{)}$ and $(\sigma_{\bar{\omega}}^2)^{ex} = 9.10^{-6}$ (dimensionless) respectively. Under the assumption of independent and uniform priors, the parameter joint posterior is written as:

$$p(\mathbf{x}, \sigma_h, \sigma_{\bar{\omega}} | \mathbf{y}_h, \mathbf{y}_{\bar{\omega}}) \propto \frac{1}{(\sigma_h \sigma_{\bar{\omega}})^N} \exp \left\{ -\frac{1}{2} \left(\frac{SS_h(\mathbf{x})}{\sigma_h^2} + \frac{SS_{\bar{\omega}}(\mathbf{x})}{\sigma_{\bar{\omega}}^2} \right) \right\}, \quad (13)$$

where SS_h and $SS_{\bar{\omega}}$ are the sum of square errors of the pressure head and average water content, respectively. The random vector $\mathbf{x} = (k_s, \omega_r, \omega_s, \alpha, n, \lambda)$ contains the soil hydraulic parameters. For the MCPD approach, maximizing the joint posterior pdf amounts to minimize the following weighted sum of squares $WSS(\mathbf{x}) = \left(\frac{SS_h(\mathbf{x})}{(\sigma_h^{\text{MAP}})^2} + \frac{SS_{\bar{\omega}}(\mathbf{x})}{(\sigma_{\bar{\omega}}^{\text{MAP}})^2} \right)$, with $(\sigma_h^{\text{MAP}})^2 = \frac{SS_h(\mathbf{x}^{\text{MAP}})}{N}$, $(\sigma_{\bar{\omega}}^{\text{MAP}})^2 = \frac{SS_{\bar{\omega}}(\mathbf{x}^{\text{MAP}})}{N}$ and $N = 481$.

The MAP estimate of the hydraulic parameters and error variances is based on the following algorithm:

1. Set $\sigma_h^{\text{MAP}} = \sigma_h^* = 1$, $\sigma_{\bar{\omega}}^{\text{MAP}} = \sigma_{\bar{\omega}}^* = 1$.
2. Find the current MAP estimate \mathbf{x}^{MAP} by minimizing the weighted sum of squares, $\mathbf{x}^{\text{MAP}} = \underset{\mathbf{x}}{\text{argmin}} WSS(\mathbf{x})$
3. Update the error variances, $(\sigma_h^{\text{MAP}})^2 = \frac{SS_h(\mathbf{x}^{\text{MAP}})}{N}$ and $(\sigma_{\bar{\omega}}^{\text{MAP}})^2 = \frac{SS_{\bar{\omega}}(\mathbf{x}^{\text{MAP}})}{N}$.
4. If $\sigma_h^{\text{MAP}} \approx \sigma_h^*$ and $\sigma_{\bar{\omega}}^{\text{MAP}} \approx \sigma_{\bar{\omega}}^*$, then stop. Otherwise set $\sigma_h^* = \sigma_h^{\text{MAP}}$ and $\sigma_{\bar{\omega}}^* = \sigma_{\bar{\omega}}^{\text{MAP}}$ and resume from 2.

Step 2 is performed with the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) which requires an initial solution (starting point). With the MCPD approach, the search of the MAP estimate is performed $N_{\text{init}} = 20$ times with different initial solutions.

With both the MCMC and MCPD samplers, eight unknowns were sought, the six hydraulic parameters $\mathbf{x} = (k_s, \omega_r, \omega_s, \alpha, n, \lambda)$ as well as the two error

variances σ_h^2 (cm²), σ_{ω}^2 (-). To speed up the MCPD evaluation, we took advantage that their estimations were independent. The eight MCPDs of interest were determined simultaneously on a multicore computer. The model was also inverted with DREAM_(ZS) for which eight chains were run simultaneously for a total maximum number of 64,000 model calls (8,000 per chain). To accelerate the convergence of the chains, the prior uncertainty range assigned to each parameter was set to the posterior plausible range obtained after the MCPD assessment (see Table 1).

Insert Figure 6 about here

5.3.2. Results and discussion

The maximal conditional posterior densities are gathered in Figure 7 as well as the parameter pairwise correlations. The bell-shaped posterior densities mean that the optimal parameter set is well identified. We note that the parameters are highly correlated, which indicates that only a small volume of the input space contains the plausible parameter sets (see also Table 1). The posterior uncertainty range of the saturated water content ω_s is particularly narrow (*i.e.* well-identified).

The saturated hydraulic conductivity k_s is positively correlated with ω_s , α , λ and negatively correlated with ω_r and n . The correlation between k_s and ω_r indicates that when fixing k_s and maximizing the conditional pdf, the estimate of ω_r is localized upon a curve (see Figure 7, row #2, column #1). This curve is slightly different when fixing ω_r and investigating the conditional estimate of k_s . Conversely, the correlation between k_s and α is so strong that the two curves coincide (row #1, column #4). We can conclude that k_s and α are virtually fully correlated. The sets of variables (ω_r, n, λ) and (ω_s, α) are also fully correlated.

The MCMC draws are depicted in the diagonal plots (Figure 7). We see that the draws are encompassed within the MCPD curves. The MCMC algorithm has virtually converged toward the true distribution after sampling $8 \times 8,000$ draws. An initial attempt with $8 \times 4,000$ draws was unsuccessful. However, it can be noticed that the MCMC sampler hardly draws values of $n > 1.3$. Because of the correlations mentioned above, this also impacts the sampling of ω_r and λ . As a consequence, Figure 7 reveals slight discrepancies between the densities of these parameters estimated with the MCPD sampler and the MCMC sampler. These results are also confirmed in Table 1, which reports the posterior uncertainty ranges. Note that the MAP estimates of the two samplers are similar and very close to the true solution \mathbf{x}^{ex} that was used to generate the data.

The MCPD sampling required about 7,500 model calls; however, because of the parallel computation, the CTU was only approximately 2,000, which corresponded to the estimate of ω_s 's MCPD. The assessment of the remaining MCPDs required far less computational efforts.

Finally, the predictive uncertainty has been assessed with the stochastic MCMC sample of size 16,000 and the 185 probabilistic MCPD draws. The 95% credible intervals are depicted in Figure 6. There is a good agreement between the two approaches. The uncertainty ranges are very narrow because many data were used for the statistical calibration (approximately one thousand).

Insert Figure 7 about here

6. Conclusions

In this work, a comparison of two sampling techniques for statistical inversion of computer models was carried out. The first technique is the well-

Parameter	k_s	ω_r	ω_s	α	n	λ
Unit	[cm/min]	[cm ³ /cm ³]	[cm ³ /cm ³]	[cm ⁻¹]	[-]	[-]
$\Omega_{\text{MCPD}}^{\text{post}}$	[0.02,0.12]	[0.05,0.18]	[0.42,0.435]	[0.008,0.011]	[1.20,1.35]	[-0.2,0.8]
$\Omega_{\text{MCMC}}^{\text{post}}$	[0.03,0.08]	[0.08,0.14]	[0.42,0.435]	[0.008,0.011]	[1.22,1.29]	[0.08,0.58]
\mathbf{x}^{ex}	0.0700	0.0900	0.4300	0.0100	1.2300	0.5000
$\mathbf{x}_{\text{MCPD}}^{\text{MAP}}$	0.0419	0.1264	0.4267	0.0090	1.2762	0.1917
$\mathbf{x}_{\text{MCMC}}^{\text{MAP}}$	0.0470	0.1176	0.4274	0.0092	1.2644	0.2683

Table 1: Parameters of the unsaturated flow model with their posterior uncertainty ranges for both MCPD and MCMC solutions. The best parameters of the MCPD and MCMC solutions are also reported. \mathbf{x}^{ex} is the set of parameters used to define the synthetic reference data of the soil drainage experiment.

known Markov Chain Monte Carlo (MCMC) sampler and the second one is a recent approach called the Maximal Conditional Posterior Distribution (MCPD) sampler.

MCMC samples stochastic draws that converge toward the desired target distribution. DREAM_(ZS), the MCMC sampler used in the present work, is a user-friendly flexible software for statistical inverse problems. It can be easily employed to infer any target distributions and does not require to modify the computer model under assessment. Several chains can be launched simultaneously to reduce the computational burden inherent to MCMC samplers but they do not evolve independently.

MCPD only samples probabilistic draws such that, for a given parameter set at a prescribed value, the other parameters maximize the conditional target distribution. Although the MCPD of a given parameter does not always match its posterior probability density function (which is inferred with MCMC), the MCPD sampler is a valuable tool for statistical inverse problems if the target distribution has a finite number of modes. For such problems, the

effectiveness of MCPD sampler enhances if the Jacobian matrix is accurately and efficiently computed which may require to modify the computer model.

In this study, the comparison between the two samplers was first carried out for two analytical distributions with collinearity and multimodality. Then they were employed to assess the posterior pdf of soil hydraulic parameters from an artificial multistep outflow experiment. For the studied problems, a good agreement is observed between the results of the two approaches. The MCPD approach was found to be less computationally demanding than DREAM_(ZS) mainly because the MCPD assessment of parameters can be performed independently and simultaneously.

Finally, it has to be mentioned that MCMC provides stochastic draws that change if the calculations are repeated. MCPD provides probabilistic draws that remain unchanged if one restarts the calculations for the same problem without changing the settings of the algorithm. This can be a problem if, posterior to the calibration, one wants to perform the uncertainty and sensitivity analysis of a model response that has not been used for the calibration. The authors are currently developing an algorithm to generate stochastic samples from MCPD draws.

Acknowledgments

This work has been funded by the French National Research Agency through the research project RESAIN (n°ANR-12-BS06-0010-02). The authors are grateful to the anonymous reviewers for their insightful comments that helped improving the manuscript. The authors are also indebted to Jasper Vrugt for the free availability of DREAM_(ZS). The MCPD sampler is written in MATLAB. The source code can be obtained upon request from the first author (mara@univ-reunion.fr).

References

- Angelikopoulos, P., Papadimitriou, C., Koumoutsakos, P., 2015. X-tmcmc: Adaptive kriging for bayesian inverse modeling. *Computer Methods in Applied Mechanics & Engineering* 289, 409–428.
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., Tu, J., 2007. A framework for validation of computer models. *Technometrics* 49, 138–154.
- Diersch, H., Perrochet, P., 1999. On the primary variable switching technique for simulating unsaturated-saturated flows. *Advances in Water Resources* 23 (3), 271–301.
- Dostert, P., Efendiev, Y., Mohanty, B., 2009. Efficient uncertainty quantification techniques in inverse problems for richards' equation using coarse-scale simulation models. *Advances in Water Resources* 32, 329–339.
- Durner, W., Iden, S. C., 2011. Extended multistep outflow method for the accurate determination of soil hydraulic properties near water saturation. *Water Resources Research* 47, W08526.
- Eching, S. O., Hopmans, J. W., Wendroth, O., 1994. Unsaturated hydraulic conductivity from transient multistep outflow and soil water pressure data. *Soil Science Society of America Journal* 58, 687–695.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–511.
- Geyer, C. J., 1992. Practical Markov chain Monte Carlo. *Statistical Science* 7 (4), 473–483.

- Grenander, U., Miller, M., 1994. Representations of knowledge in complex system. *Journal of the Royal Statistical Society, Serie B* 56 (4), 549–603.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: Efficient adaptive MCMC. *Statistics and Computing* 16 (339–354).
- Haario, H., Saksman, Tamminen, J., 2001. An adaptive Metropolis algorithm. *Bernoulli* 7, 223–242.
- Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics* 14, 377–395.
- Hastings, H., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hayek, M., Lehmann, F., Ackerer, P., 2008. Adaptive multi-scale parameterization for one-dimensional flow in unsaturated porous media. *Advances in Water Resources* 31, 28–43.
- Laloy, E., Vrugt, J. A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(zs) and high-performance computing. *Water Resources Research* 48, W01526.
- Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics* 2, 164–168.
- Lu, D., Ye, M., Hill, M. C., Poeter, E. P., Curtis, G. P., 2014. A computer program for uncertainty analysis integrating regression and bayesian methods. *Environmental Modelling and Software* 60, 45–56.
- Mara, T. A., Fajraoui, N., Younes, A., Delay, F., 2015. Inversion and uncertainty of highly parameterized models in a bayesian framework by sampling

- the maximal conditional posterior distribution of parameters. *Advances in Water Resources* 76, 1–10.
- Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11, 431–441.
- Metropolis, N.-A., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087–1091.
- Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research* 12, 513–522.
- Robert, C. P., Casella, G., 2004. *Monte Carlo statistical method*. Springer series in statistics, second edition, New York.
- ter Braak, C. J. F., 2006. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing* 16, 239–249.
- ter Braak, C. J. F., Vrugt, J., 2008. Differential evolution markov chain with snooker updater and fewer chains. *Statistics and Computing* 18 (4), 435–446.
- Tierney, L., Mira, A., 1999. Some adaptive monte carlo methods for bayesian inference. *Statistics in Medicine* 18, 2507–2515.
- van Dam, J. C., Stricker, J. N. M., Droogers, P., 1994. Inverse method to determine soil hydraulic functions from multistep outflow experiment. *Soil Science Society of America Journal* 58, 647–652.

- van Genuchten, M. T., 1980. A closed form equation for predicting the hydraulic properties of unsaturated soils. *Soil Science Society of America Journal* 44 (5), 892–898.
- Vrugt, J. A., 2016. Markov chain monte carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software* 75, 273–316.
- Vrugt, J. A., Bouten, W., 2002. Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. *Soil Science Society of America Journal* 66, 1740–1751.
- Vrugt, J. A., Gupta, H. V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39 (8), 1201.
- Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Higdon, D., Robinson, B. A., Hyman, J. M., 2009. Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation* 10, 273–290.

Figures

List of Figures

1	The MCPD refinement	32
2	Banana-shaped distributions 1	33
3	Gelman-Rubin convergence diagnostic	34
4	MCPD versus MCMC: Multimodal case	35
5	Chain evolution	36
6	Data used for the soil hydraulic parameter identification	37
7	MCPDs of the unsaturated soil hydraulic parameters	38

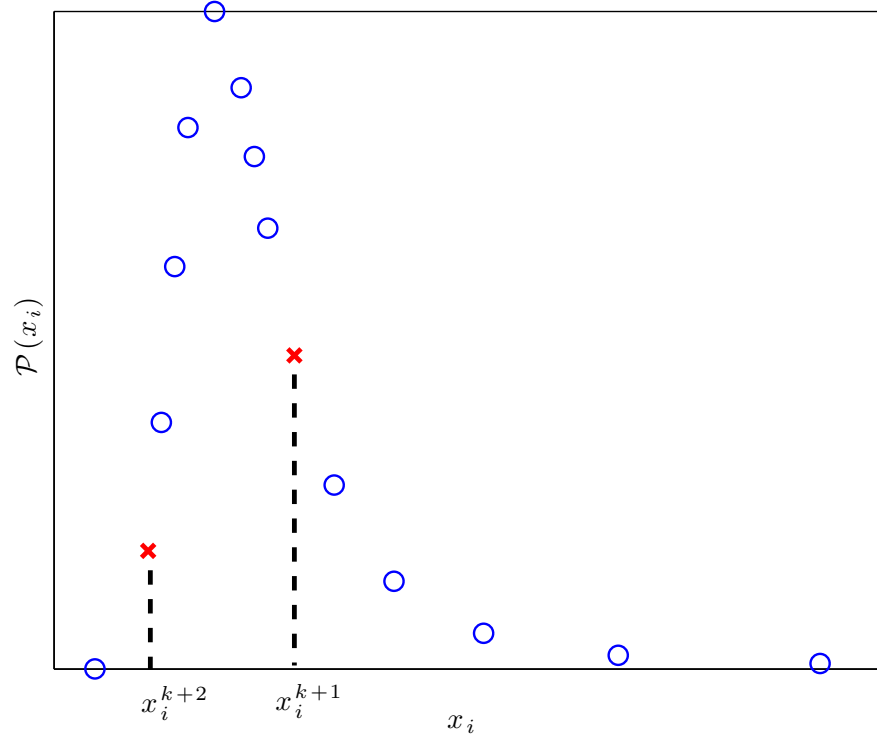


Figure 1: Refinement of the MCPD in the third part of the algorithm. The circle plots represent the MCPD assessment of x_i after the second part of the calculations. The crosses indicate the next draws selected in the third part of the algorithm.

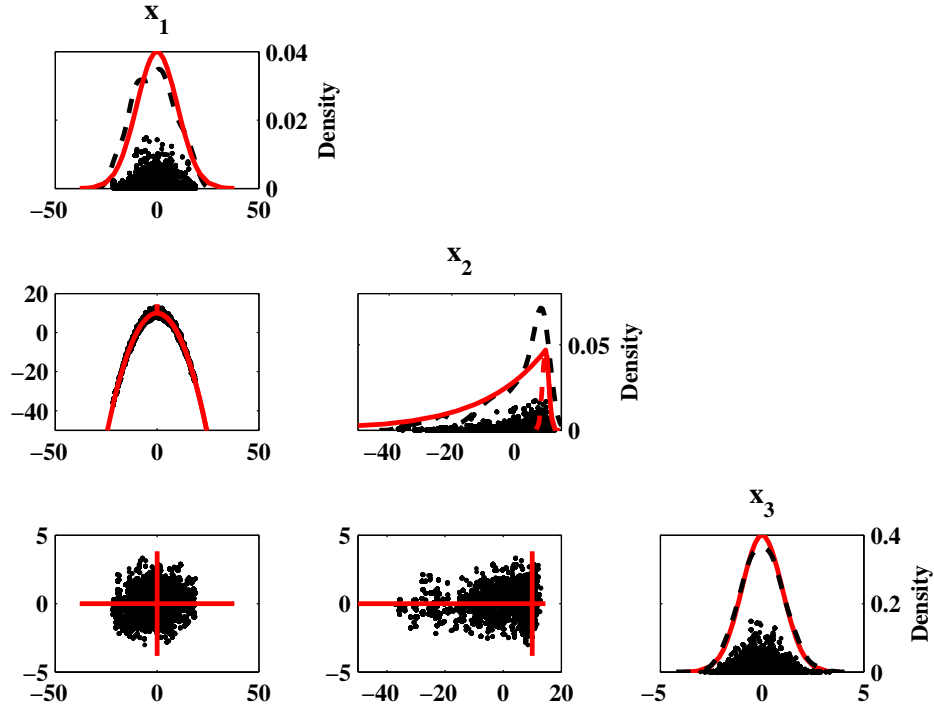


Figure 2: Assessment of the probability densities of the parameters in the twisted-Gaussian function. The continuous and broken red lines represent the MCPD draws while the black dots and broken lines represent the MCMC draws. The diagonal plots represent the estimated posterior densities. The off-diagonal plots depict the pairwise correlations. Note the banana-shaped relationship between (x_1, x_2) (row #2, column #1).

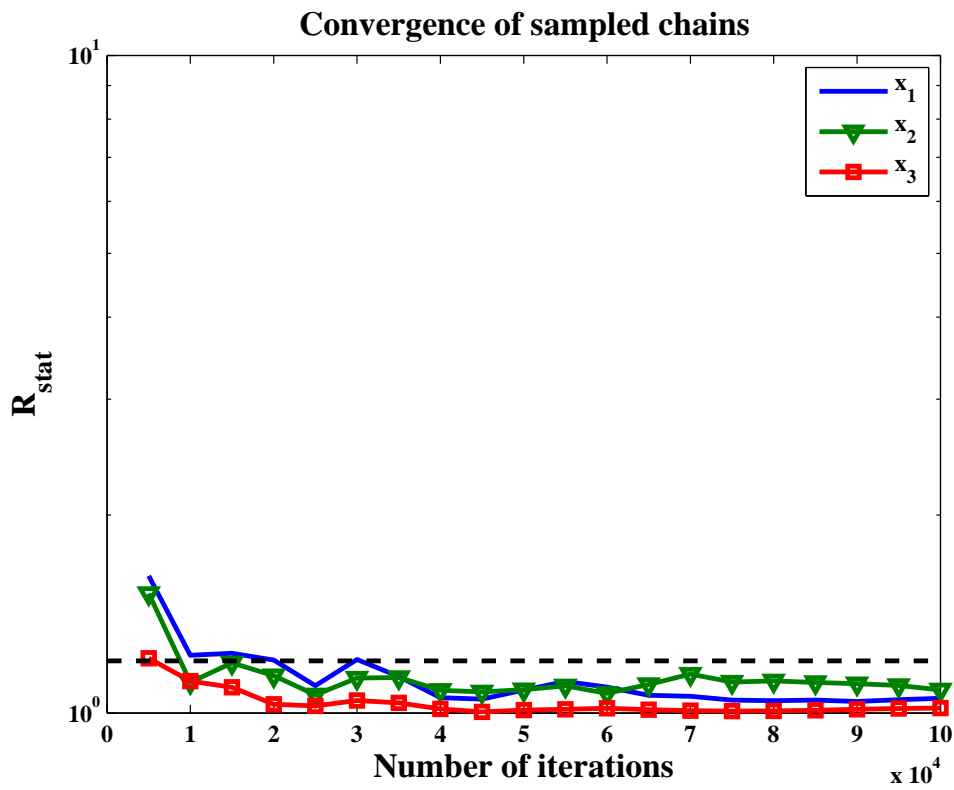


Figure 3: Evolution of the Gelman-Rubin \hat{R} -statistics for the convergence diagnostic of the first three parameters. The convergence criterion is achieved if the chains reach the threshold in broken-line ($\hat{R} \leq 1.2$).

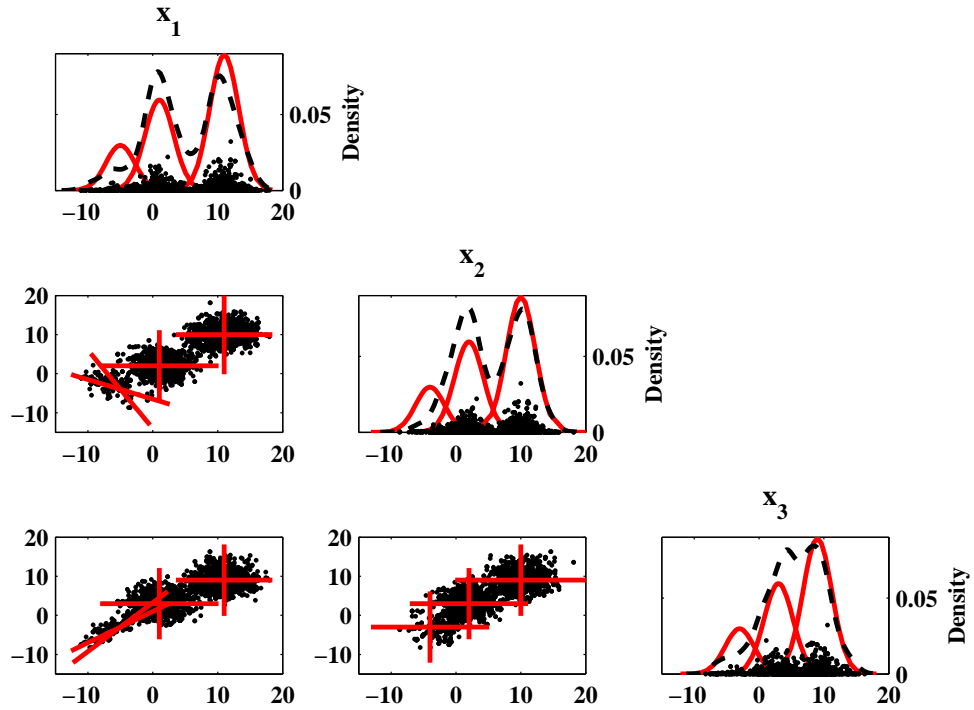


Figure 4: Assessment of the probability density of the parameters in the multimodal function. For explanations, see the label in Figure 2 and the body text.

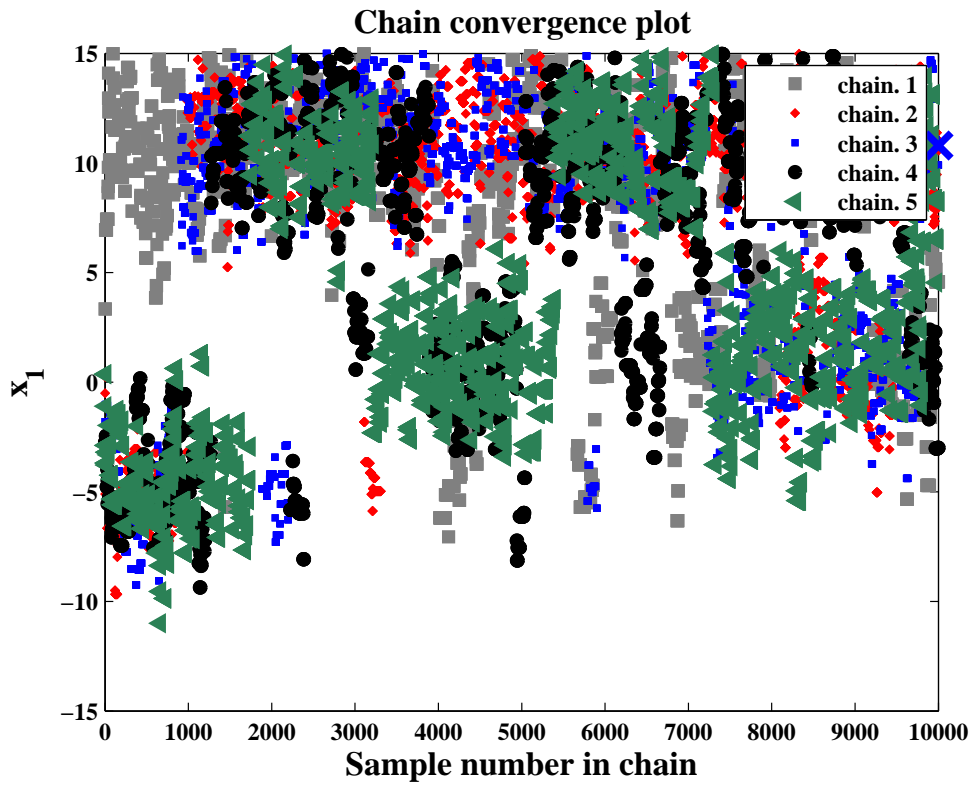


Figure 5: Evolution of the MCMC draws of x_1 in the first five chains (over eleven parallel chains). The first 3,000 draws are located around $x_1 = -5$ and $x_1 = 11$. Then, the sampler localizes the last local optimum $x_1 = 1$ and samples around the three local optima.

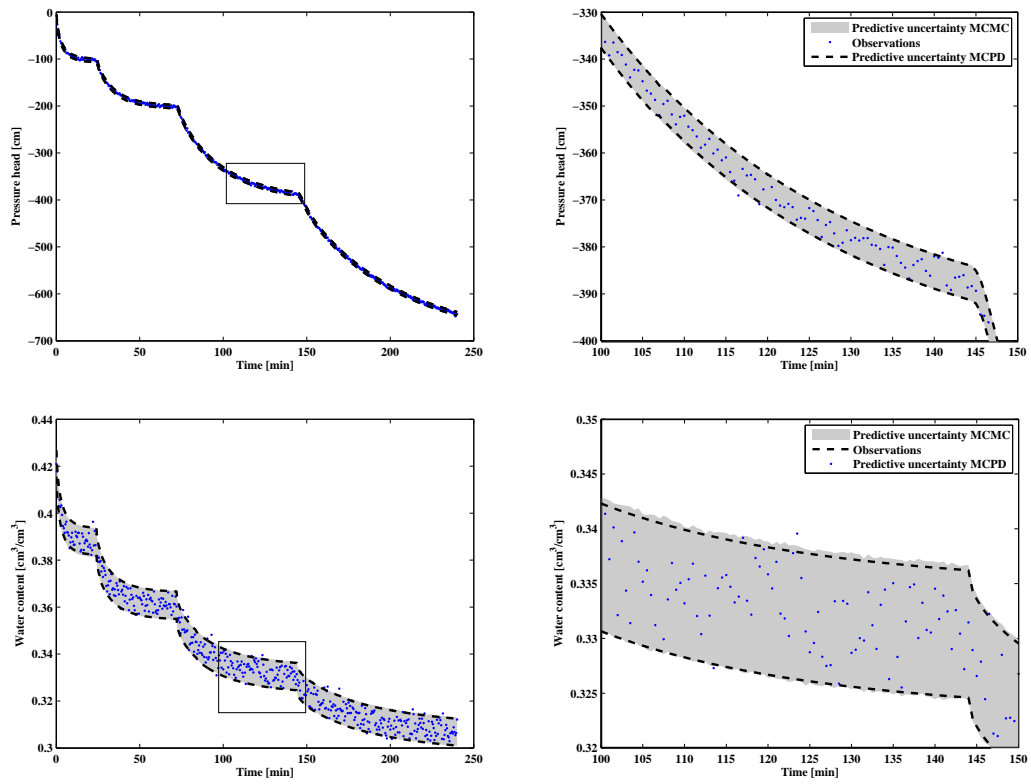


Figure 6: The data used for soil hydraulic parameter identification. The narrow predictive uncertainty ranges assessed with the MCPD and MCMC samples respectively indicate that the model is satisfactorily calibrated.

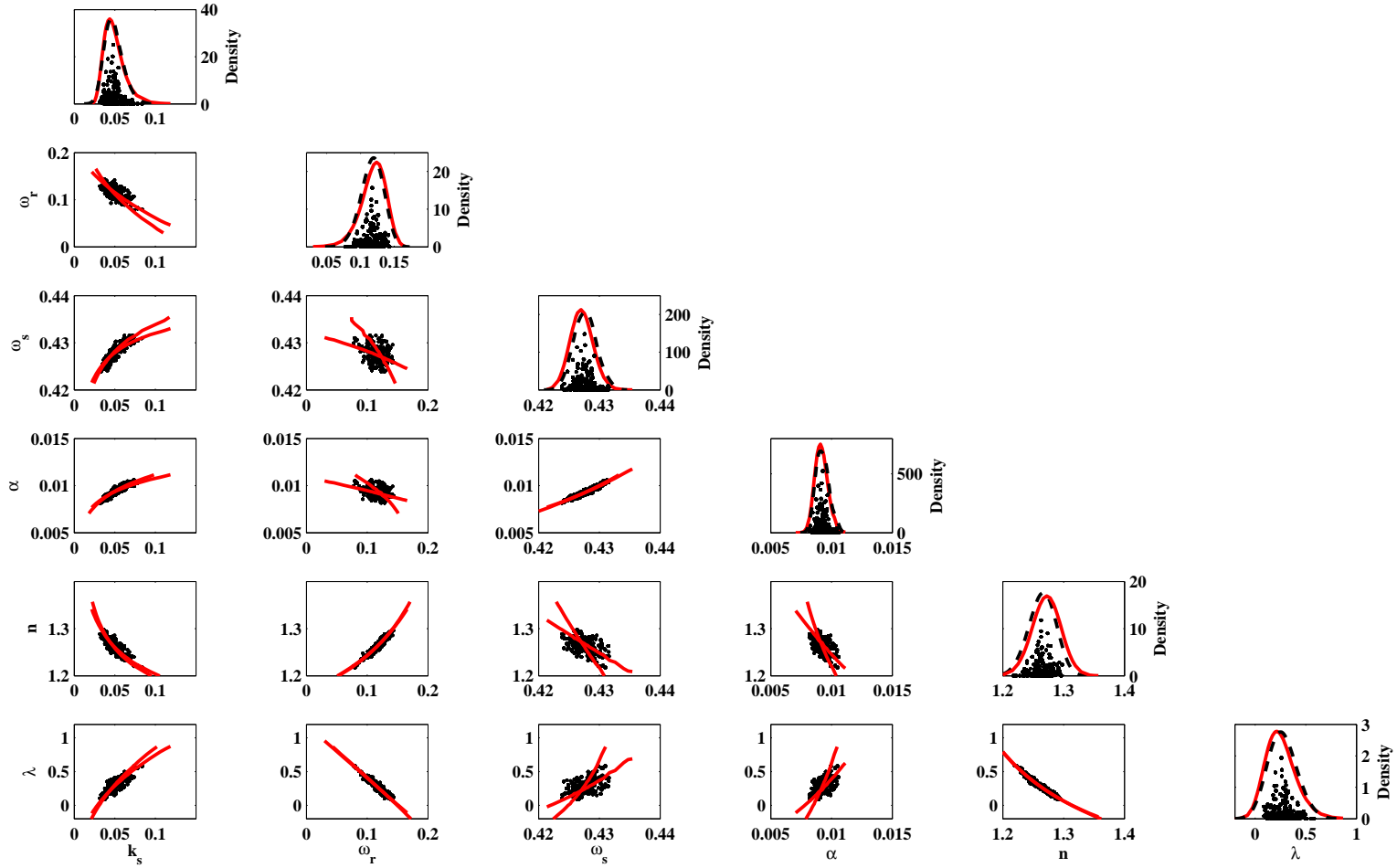


Figure 7: Densities of the soil hydraulic parameters estimated with the MCPD approach (continuous red curves) and with DREAM_(ZS) (scatterplots+estimated densities in broken lines) are depicted on the diagonal plots. The off-diagonal curves represent the pairwise correlations between the soil hydraulic parameters. We note the strong correlations amongst the parameters, notably between (ω_r, n, λ) and (ω_s, α) respectively.