



**HAL**  
open science

# Inversion and uncertainty of highly parameterized models in a Bayesian framework by sampling the maximal conditional posterior distribution of parameters

Thierry A. Mara, Noura Fajraoui, Anis Younes, Frederick Delay

## ► To cite this version:

Thierry A. Mara, Noura Fajraoui, Anis Younes, Frederick Delay. Inversion and uncertainty of highly parameterized models in a Bayesian framework by sampling the maximal conditional posterior distribution of parameters . *Advances in Water Resources*, 2015, 76, pp.1 - 10. 10.1016/j.advwatres.2014.11.013 . hal-01098506

**HAL Id: hal-01098506**

**<https://hal.science/hal-01098506>**

Submitted on 26 Dec 2014

**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.

# Inversion and uncertainty of highly parameterized models in a Bayesian framework by sampling the maximal conditional posterior distribution of parameters\*

Thierry A. Mara<sup>a,\*</sup>, Noura Fajraoui<sup>b</sup>, Anis Younes<sup>b</sup>, Frederick Delay<sup>b</sup>

<sup>a</sup>*PIMENT, EA 4518, Université de La Réunion, FST, 15 Avenue René Cassin, 97715 Saint-Denis, Réunion*

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

---

## Abstract

We introduce the concept of Maximal Conditional Posterior Distribution (MCPD) to assess the uncertainty of model parameters in a Bayesian framework. Although, Markov Chains Monte Carlo (MCMC) methods are particularly suited for this task, they become challenging with highly parameterized nonlinear models. The MCPD represents the conditional probability distribution function of a given parameter knowing that the other parameters maximize the conditional posterior density function. Unlike MCMC which accepts or rejects solutions sampled in the parameter space, MCPD is calculated through several optimization processes. Model inversion using MCPD algorithm is particularly useful for highly parameterized problems because calculations are independent. Consequently, they can be evaluated simultaneously with a multi-core computer. In the present work, the MCPD approach is applied to invert a 2D stochastic groundwater flow problem where

---

\*Corresponding author: mara@univ-reunion.fr

\*Published in *Advances in Water Resources*, Vol. 76, pp. 1-10, 2014, doi:10.1016/j.advwatres.2014.11.013

the log-transmissivity field of the medium is inferred from scarce and noisy data. For this purpose, the stochastic field is expanded onto a set of orthogonal functions using a Karhunen-Loève (KL) transformation. Though the prior guess on the stochastic structure (covariance) of the transmissivity field is erroneous, the MCPD inference of the KL coefficients is able to extract relevant inverse solutions.

*Keywords:* Inverse modeling, Bayesian parameter estimation, Model parameter identification, Highly parameterized model, Heterogeneous transmissivity field, Karhunen-Loève expansion.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Bayesian inference</b>	<b>6</b>
<b>3</b>	<b>Maximal conditional posterior distribution</b>	<b>7</b>
3.1	The concept . . . . .	7
3.2	MCPD assessment . . . . .	9
3.3	Predictive density of an MCPD sample . . . . .	11
<b>4</b>	<b>Inference of a multimodal distribution with MCPDs</b>	<b>12</b>
<b>5</b>	<b>Inverting a stochastic field with MCPDs</b>	<b>15</b>
5.1	Karhunen-Loève expansion . . . . .	15
5.2	Problem statement . . . . .	17
5.3	Posterior distribution . . . . .	18
5.4	Results and discussion . . . . .	21
<b>6</b>	<b>Conclusions</b>	<b>25</b>

## 1. Introduction

Models are tools on which environmental risk-assessment and decision-making strategies can rely, provided it is proved that the models are relevant to the problem under investigation. This relevance can be addressed by facing a model prediction to observation data knowing that the whole procedure also requires assigning model parameter values. Some parameters can be directly measured while some others ought to be indirectly estimated by comparing model predictions with observations. The present work addresses the issue of parameter identification for highly parameterized models. The notion of identification encompasses seeking the parameter values and assessing the uncertainty on parameters and on model predictions.

During the past two decades, the increasing power of computers was conducive to emphasize and promote the so-called Bayesian parameter estimation techniques. In essence, the Bayesian framework leads to the definition of the parameter joint posterior probability density function (pdf), for instance inferred by means of Markov Chain Monte Carlo (MCMC) samplings ([1–4]). The notion of posterior pdf is associated with the fact that the parameter’s pdf is conditioned both on plausible (prior) parameter values and on observation data. MCMC provides draws directly sampled from the posterior pdf which leads to exploration of the plausible areas in the parameter space. The Bayesian estimation using MCMC has been subject to many developments and improvements during the last decade (e.g. [5–8] among others). However, MCMC samplers remain computationally expensive because many draws are rejected by the statistical test embedded in the sampler. Furthermore, with MCMC, the parameters marginal posterior distributions cannot be investigated independently. Recently, several strategies have been proposed to increase MCMC efficiency (see [9–13]).

In the present work we propose a new method, partly grounded in optimization techniques, to cope with the identification of model parameters. The first step of this approach is to seek all the probable local optima of the joint posterior pdf of the whole set of parameters (including the maximum a posteriori estimate). Next, several maximizations of the conditional pdf are performed for different prescribed values of one selected parameter. The values assigned to this parameter are picked from a range around its optimal value(s). The value of the other parameters is investigated by maximizing the conditional pdf. This provides what we call the Maximal Conditional Posterior Distribution (MCPD) of the selected parameter. It actually corresponds to a discrete approximation of the pdf of a single parameter conditioned on data such that the conditional pdf is maximized.

The MCPD returns information about the model parameter values supported by the data and any correlations between parameters. The MCPD sample also allows uncertainty bounds to be assigned to the model predictions. The main advantage of the approach is that MCPD inferences for different parameters are independent and can be evaluated simultaneously by easily distributing the calculations over a multi-core computer (or several computers). This feature drastically decreases the computation time and makes the inversion of highly parameterized problems feasible.

The main topics addressed in the present paper are organized as follows. A short outline on inverse modeling within a Bayesian framework is proposed in Section 2, and then followed by the details on the MCPD sampling in Section 3. The first exercise testing the MCPD approach is proposed in Section 4 and addresses the ability of the sampler to retrieve a multimodal probability density function. The second test in Section 5 applies the MPCD approach to identify the Karhunen-Loève expansion ([14]) of a stochastic transmissivity

field for a two-dimensional steady-state groundwater flow problem.

## 2. Bayesian inference

In inverse modeling, the parameter set (of size  $s$ )  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_s\}$  of a given model is estimated from a set of observation data  $\mathbf{d}$ . In the following, we assume that the model does not suffer from misconceptions. The model is therefore supposed to be exact regarding the processes and the system that it mimics. However, observation data remain uncertain (random variables) making the model parameters to be also random and characterized by a joint probability density function  $p(\boldsymbol{\theta})$ . We denote by  $\Omega_i$  the probable prior uncertainty range of  $\theta_i$ . In a Bayesian framework, the parameter joint posterior pdf is defined by

$$p(\boldsymbol{\theta}|\mathbf{d}) = \frac{p(\mathbf{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d})} \quad (1)$$

where  $p(\mathbf{d})$  is a scaling factor called evidence,  $p(\boldsymbol{\theta})$  is the prior density corresponding to a first guess on parameters before collecting the observations, while  $p(\mathbf{d}|\boldsymbol{\theta})$  is called the likelihood function measuring how well the model describes the data.

The parameter set that maximizes Eq. (1)

$$\boldsymbol{\theta}^{MAP} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{d}) \quad (2)$$

is called the maximum a posteriori estimate. It is the most probable parameter set given our knowledge about the system (i.e. the data  $\mathbf{d}$  and the prior pdf of the parameters  $p(\boldsymbol{\theta})$ ) and it is sought by appropriate optimization algorithms (e.g., descent methods, evolutionary algorithms, etc...). Unfortunately, finding  $\boldsymbol{\theta}^{MAP}$  does not allow to (fully) characterize the posterior uncertainty of the parameters (except for linear models, see [15]). This uncertainty should be assessed by calculating the marginal posterior density for

each parameter, defined as follows

$$p(\theta_i|\mathbf{d}) = \int p(\theta_i, \boldsymbol{\theta}_{-i}|\mathbf{d})d\boldsymbol{\theta}_{-i}, \quad \forall i = 1, \dots, s \quad (3)$$

where  $\boldsymbol{\theta}_{-i}$  represents the vector of parameters  $\boldsymbol{\theta}$  without  $\theta_i$ . The integral in (3) can be approximated by a multidimensional quadrature method or by a sampling-based method such as the Markov Chain Monte Carlo (MCMC). Nevertheless, the computational effort can be prohibitive and sometimes unaffordable for problems with a large number of parameters.

In the present work, we propose an optimization-based method in order to assess the parameter uncertainty for models post-conditioned on available observation data. For this purpose, we introduce the concept of maximal conditional posterior distribution. One could raise that relying on an optimization-based method will require solving many problems, as is classical with standard inversion techniques when obtaining a large set of solutions is contemplated. As shown hereafter, the maximal conditional posterior distribution has some specific features diminishing the calculation loads.

### 3. Maximal conditional posterior distribution

#### 3.1. The concept

We define the maximal conditional posterior distribution (MCPD) of  $\theta_i$  as

$$\mathcal{P}_i(\theta_i) = \max_{\boldsymbol{\theta}_{-i}} (p(\boldsymbol{\theta}_{-i}|\mathbf{d}, \theta_i)) \times p(\theta_i|\mathbf{d}) \quad (4)$$

$\mathcal{P}_i(\theta_i)$  is interpreted as the posterior probability function that maximizes the conditional posterior distribution  $p(\boldsymbol{\theta}_{-i}|\mathbf{d}, \theta_i)$  and encompasses the MAP probability (i.e.  $\mathcal{P}_i(\theta_i^{MAP}) = p(\boldsymbol{\theta}^{MAP}|\mathbf{d})$ ). By using the Bayes theorem, one can write,  $\max (p(\boldsymbol{\theta}_{-i}|\mathbf{d}, \theta_i)) \times p(\theta_i|\mathbf{d}) = \max (p(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{d}) \times p(\boldsymbol{\theta}_{-i}|\mathbf{d}))$ . Therefore, the MCPD in (4) can also be viewed as the distribution of the



parameter  $\theta_i$ , knowing that the other parameters  $\boldsymbol{\theta}_{-i}$  are at their optimal values. The MCPD of  $\theta_i$  is assessed in a discrete form by sampling Eq. (4). A parameter  $\theta_i$  is frozen at a prescribed value and the other parameters  $\boldsymbol{\theta}_{-i}$  are optimized to find (according to the Bayesian definition) the maximal probability of these parameters. Changing the prescribed value of  $\theta_i$  allows scanning the distribution of  $\theta_i$ . In practice, the sampled values of  $\theta_i$  (denoted below  $\theta_i^*$ ) are picked around the MAP estimate  $\theta_i^{MAP}$  (estimated beforehand) within its prior uncertainty range  $\Omega_i$  (see Fig. 1). This gives,

$$\boldsymbol{\theta}_{-i}^* = \arg \max_{\boldsymbol{\theta}_{-i}} p(\boldsymbol{\theta}_{-i} | \mathbf{d}, \theta_i = \theta_i^*) \quad (5)$$

$$\mathcal{P}_i(\theta_i^*) = p(\boldsymbol{\theta}_{-i}^* | \mathbf{d}, \theta_i^*) \times p(\theta_i^* | \mathbf{d}) = p(\boldsymbol{\theta}^* | \mathbf{d}) \quad (6)$$

On the one hand, if  $\theta_i$  is globally identifiable, one expects that the further the parameter value is from  $\theta_i^{MAP}$ , the more  $\mathcal{P}_i(\theta_i^*) < \mathcal{P}_i(\theta_i^{MAP})$ . On the other hand, if  $\theta_i$  is not identifiable, varying its sampled values  $\theta_i^*$ , will not change the joint posterior distribution in (1). Then,  $\max(p(\boldsymbol{\theta}_{-i} | \mathbf{d}, \theta_i)) = p(\boldsymbol{\theta}_{-i}^{MAP} | \mathbf{d})$  and the MCPD of  $\theta_i$  will be equal to its marginal prior distribution (i.e.  $\mathcal{P}_i(\theta_i) \propto p(\theta_i)$ , see Eq. (4)). This is the case for instance of the unimodal multi-Gaussian probability density function.

In the event of multimodality of  $p(\boldsymbol{\theta} | \mathbf{d})$ ,  $\boldsymbol{\theta}^{MAP}$  is not unique or its search is hampered by the existence of local maxima. It is advisable to first acquire all the possible maxima of  $p(\boldsymbol{\theta} | \mathbf{d})$  by starting optimization processes from different locations in the parameter space (a multi-start procedure). Note that these prior searches do not ensure that all maxima will be identified. When the local maxima are known, the MCPD calculation principle evoked above is generalized by sampling  $\theta_i^*$  in all the subareas of the parameter space enclosing the local maxima. Finally, note that, by definition, the MCPD  $\mathcal{P}_i(\theta_i)$  and the marginal posterior density  $p(\theta_i | \mathbf{d})$  are different. However,

these densities would be exactly the same whenever  $\max(p(\boldsymbol{\theta}_{-i}|\mathbf{d}, \boldsymbol{\theta}_i))$  is constant  $\forall \boldsymbol{\theta}_i \in \Omega_i$ .

It can be questioned on the relevance of the MCPD if the latter is not the marginal posterior density. We remind that the aim of the MCPD is to evaluate uncertainties on both the parameters and the model predictions given some observation data. The key point is less to know whether one can approach the marginal posterior density rather than providing uncertainties for valuable solutions to the inverse problem. As told above, by construction an MCPD samples the distribution of a parameter  $\boldsymbol{\theta}_i$  knowing that the others are optimal. This new way brought by the MCPD to envision the uncertainty associated with a parameter seems to be an interesting alternative to more classical definitions. One can mention however that a Markov chain Monte Carlo sampler tries to draw all the probable solutions to the inverse problem while only some of these solutions are inferred by the MCPD sampling. In some situations, this can represent a drawback of the MCPD sampling.

### 3.2. MCPD assessment

For the sake of clarity, the algorithm inferring the MCPD assumes that the posterior density  $p(\boldsymbol{\theta}|\mathbf{d})$  has only one mode. The occurrence of multimodal densities is just an adaptation of the procedure depicted below. The procedure starts by seeking the maximum a posteriori  $\boldsymbol{\theta}^{MAP}$  (Eq. (2)). Then, the algorithm sketched by the flowchart in Fig. 2 is launched. Calculation of MCPD for each parameter takes place in two stages. The first stage identifies the relevant range within which the parameter  $\boldsymbol{\theta}_i$  will be made to vary. For this purpose, we define a large sampling step, for  $\boldsymbol{\theta}_i$ , e.g.  $\Delta \times \boldsymbol{\theta}_i^{MAP}$ , with  $\Delta = 0.5$ . Then,  $\boldsymbol{\theta}_i$  is successively set to  $\boldsymbol{\theta}_i^{\pm k} = \boldsymbol{\theta}_i^{MAP}(1 \pm k\Delta)$ ,  $k = 1, 2, 3, \dots$  and the associated optimization in Eq. (5) is solved. Increments of  $k$  are stopped when:

1. the two current probability values  $p(\boldsymbol{\theta}^{+k}|\mathbf{d})$  and  $p(\boldsymbol{\theta}^{-k}|\mathbf{d})$  decrease below a prescribed threshold which may be expressed thus:  $p(\boldsymbol{\theta}^{\pm k}|\mathbf{d}) < p(\boldsymbol{\theta}^{MAP}|\mathbf{d})/100$ ; or
2. the sampled values  $\theta_i^{\pm k}$  are beyond  $\Omega_i$ , i.e. the prescribed prior uncertainty range on  $\theta_i$ .

Usually, the first stopping criterion is reached before the second one, thus leading to a narrower sampling range of  $\theta_i$  compared to its prior uncertainty range  $\Omega_i$ . The second stage of the algorithm resamples the values of  $\theta_i$  to increase the refinement of the discrete MCPD. Before proceeding, the results from the first sampling step are re-ranked by increasing values of  $\theta_i$ . We denote  $\{\boldsymbol{\theta}^{k_i}, \mathcal{P}_i(\boldsymbol{\theta}_i^{k_i}) = p(\boldsymbol{\theta}^{k_i}|\mathbf{d}), k_i = 1, \dots, n_i\}$  the re-ranked first sample with  $\theta_i^{k_i} < \theta_i^{k_i+1}$ . We seek in this sample the interval where the difference  $\mathcal{P}_i(\theta_i^{k_i+1}) - \mathcal{P}_i(\theta_i^{k_i})$  is maximal, i.e., we seek the index  $k_m$  verifying

$$k_m = \arg \max_{k_i} |\mathcal{P}_i(\theta_i^{k_i+1}) - \mathcal{P}_i(\theta_i^{k_i})| \quad (7)$$

Then, the optimization in Eq. (5) is solved for  $\theta_i^* = \frac{\theta_i^{k_m+1} + \theta_i^{k_m}}{2}$  and the pair  $(\theta_i^*, \mathcal{P}_i(\theta_i^*))$  joins the set  $\{\boldsymbol{\theta}^{k_i}, \mathcal{P}_i(\theta_i^{k_i})\}$ . The latter is then sorted again by increasing value of  $\theta_i$  and the search of the index  $k_m$  is resumed. This procedure makes that one resamples  $\theta_i$  and amends the set  $\{\boldsymbol{\theta}^{k_i}, \mathcal{P}_i(\theta_i^{k_i})\}$  progressively (iteratively) at locations in the set where the MCPD is the most coarsely discretized. Usually, a few iterations on  $k_m$  (i.e.  $N_{it} = 10$ ) are enough to obtain a good discrete depiction of the whole MCPD of  $\theta_i$ .

In the case of multimodal densities  $p(\boldsymbol{\theta}|\mathbf{d})$ , the algorithm is adapted so that it repeats the first and second steps of sampling around each "optimal"  $\theta^{opt}$  value (see the example in Section 4).

At this stage, it is interesting to note that the optimization procedures to identify optimal vectors  $\boldsymbol{\theta}_{-i}$  for prescribed values of  $\theta_i$  are completely

independent of the optimizations identifying  $\boldsymbol{\theta}_{-j}$  for prescribed values of  $\theta_j$ . Therefore, the MCPD samplings for each parameter of the investigated problem are independent and can be handled easily with parallel computing streams. There can be as many sessions as the number of parameters, which strongly reduces the total computation time. Obviously, other inverse techniques can be parallelized, for example, the Markov Chain Monte Carlo (see [16], [17], [18]). In this case, several chains are launched in parallel to explore the parameter space. The chains will generate independent subpopulations of solutions and exchange good individuals between the subpopulations to accelerate the convergence (the benefit of some emigration between subpopulations). The parallel calculations are no longer independent and a master computer is needed to analyze the independent subpopulations and generate new ones. In the end, the ease of parallelization brought by the MCPD inference is suited to the inversion of highly parameterized problems (see the example in Section 5).

### 3.3. Predictive density of an MCPD sample

Let us denote by  $n = \sum_{i=1}^s n_i$  the total number of MCPD draws. A specific MCPD sample  $(\boldsymbol{\theta}^k, \mathcal{P}_i(\boldsymbol{\theta}_i^k))$ ,  $k = 1, \dots, n$  corresponds to a point location on the hypersurface described by the parameter joint probability density  $p(\boldsymbol{\theta}|\mathbf{d})$ . The density function of the model predictions conditioned on observed data  $\mathbf{d}$  allows assigning uncertainty bounds to the prediction of a new observation  $d^*$ . If we assume that  $d^*$  is independent of  $\mathbf{d}$  conditional on  $\boldsymbol{\theta}$  (i.e.  $p(d^*|\mathbf{d}, \boldsymbol{\theta}) = p(d^*|\boldsymbol{\theta})$ , see [19]), we can write

$$p(d^*|\mathbf{d}) = \int p(d^*|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{d})d\boldsymbol{\theta} \quad (8)$$

Given that,  $\mathcal{P}_i(\boldsymbol{\theta}_i^k) = p(\boldsymbol{\theta}^k|\mathbf{d})$ , Eq. (8) is simply approximated by

$$\hat{p}(d^*|\mathbf{d}) = \frac{\sum_{k=1}^n \mathcal{P}_i(\boldsymbol{\theta}_i^k)p(d^*|\boldsymbol{\theta}^k)}{\sum_{k=1}^n \mathcal{P}_i(\boldsymbol{\theta}_i^k)} \quad (9)$$

where  $p(d^*|\boldsymbol{\theta}^k)$  is the likelihood function defined in Eq. (1) evaluated at  $d^*$  and conditioned onto  $\boldsymbol{\theta} = \boldsymbol{\theta}^k$ .

#### 4. Inference of a multimodal distribution with MCPDs

The first case study deals with the ability of MCPD samplings to retrieve a three-modal probability density function with connected and disconnected modes. The test function to be retrieved is inspired from a case study proposed by Laloy and Vrugt who state that inferring a density function with disconnected modes by means of MCMC is very challenging ([20]). In the present work, the density function also encloses modes that differ for each parameter. The density function to retrieve has 25 parameters and is the sum of three multiGaussian density functions expressed as

$$p(\boldsymbol{\theta}) = \frac{1}{6}\mathcal{N}(\boldsymbol{\mu}_1, 5\mathbf{C}) + \frac{2}{6}\mathcal{N}(\boldsymbol{\mu}_2, 5\mathbf{I}_{25}) + \frac{3}{6}\mathcal{N}(\boldsymbol{\mu}_3, 5\mathbf{I}_{25}) \quad (10)$$

where  $\mathcal{N}(\boldsymbol{\mu}_i, \mathbf{I}_{25})$  is the multiGaussian distribution of mean vector  $\boldsymbol{\mu}_i$ .  $\mathbf{I}_{25}$  is the 25-dimensional identity matrix which indicates that the parameters  $(\theta_1, \dots, \theta_{25})$  are independent in the second and third Gaussian distributions in (10).  $\mathbf{C}$  is a correlation matrix with null off-diagonal elements except for  $C_{1,11} = C_{11,1} = -0.5$  and  $C_{1,13} = C_{13,1} = 0.8$ . These non-null terms impose, for the first Gaussian distribution in Eq. (10), a negative correlation between  $\theta_1$  and  $\theta_{11}$  and a strong positive correlation between  $\theta_1$  and  $\theta_{13}$ . The three modes of each parameter are grouped in the vectors of means  $\boldsymbol{\mu}_1 = [-12, \dots, 12]$ ,  $\boldsymbol{\mu}_2 = [1, \dots, 25]$  and  $\boldsymbol{\mu}_3 = [25, \dots, 1]$ . Thus,  $\theta_{13}$  has two modes located at  $\theta_{13} = 13$ . To compute the MCPDs, and specifically to seek optimal vector  $\boldsymbol{\theta}_{-i}^*$  knowing the value of  $\theta_i^*$ , we use the MATLAB optimization toolbox, especially the *fminunc.m* function based on the so-called trust-region method. This function minimizes  $-\log(p(\boldsymbol{\theta}))$  and the

convergence velocity is accelerated by providing the values of the derivatives of  $-\log(p(\boldsymbol{\theta}))$  with respect to the  $\boldsymbol{\theta}$  components.

The first hurdle in this exercise is to identify the 3 local maxima of the density function. For all parameters, the maximal mean values are located in the third multiGaussian distribution with  $\boldsymbol{\mu}_3$  as the vector of means. For this purpose we ran the optimization procedure twenty times, using initial solutions uniformly sampled from the parameter space. This calculation required about 6 500 evaluations of the density function (Eq. (10)). Subsequently, MCPD sampling was carried out in the vicinity of each optimum requiring an additional computation effort of about 3 000 runs for  $N_{it} = 20$  iterations of MCPDs refinements (see Section 3).

In Fig. 3 we compare the maximal conditional posterior density of some parameters to their marginal posterior density. The MCPDs were computed numerically while the marginal posterior densities were assessed analytically by computing  $p(\theta_i) = \int p(\boldsymbol{\theta})d\boldsymbol{\theta}_{-i}$ . Interestingly, when the modes are disconnected the two densities are equal (e.g.  $\theta_1$ ) whereas they can be very different when the modes are superimposed (e.g.  $\theta_{13}$ ). This is due to the fact that the marginal posterior pdf is an integral making that the superimposed modes are summed-up. Note that because the MCPDs are calculated by several optimization procedures in the vicinity of each mode, the superimposed modes produce overlapping MCPD curves.

It is noticeable that with only a few point estimates of the MCPD, the densities obtained are accurate enough (e.g. with only  $N_{it} = 10$  refinement iterations, not reported in Fig. 3). The "off-diagonal" scatterplots in Fig. 3 for a column  $i$  and a row  $j$  correspond to the pairwise MCPD draws  $(\theta_i^{k_i}, \theta_j^{k_i})$  and  $(\theta_i^{k_j}, \theta_j^{k_j})$ . The first one corresponds to the optimal sought value of the parameter  $\theta_j$  for sampled (prescribed) values of  $\theta_i$  and the second one to

optimal values of  $\theta_i$  for sampled  $\theta_j$ . The first observation is that one can find several distinct values of a parameter  $\theta_i$  for the same draws of  $\theta_j$ . This is specific to the targeted pdf with overlapping modes. For example in Fig. 3,  $\theta_{13}$  has its three modes in  $(0, 13, 13)$ , i.e. two overlapping modes, whereas modes of  $\theta_1$  at  $(-12, 1, 25)$  do not overlap. When seeking the MCPD of  $\theta_{13}$ , one identifies two local optima ( $\theta_1 = 1, \theta_{13} = 13$ ) and ( $\theta_1 = 25, \theta_{13} = 13$ ), i.e., several distinct values of  $\theta_1$  (around 1 and 25, respectively) are found for the same draws of  $\theta_{13}$  around the value 13.

The scatterplots also prove that the MCPD sampler is able to retrieve the correlation structure of the pdf. One can easily check that, in the vicinity of the first mode (for which  $\theta_1 = 1, \theta_{11} = 11$  and  $\theta_{13} = 13$ ), a negative correlation between  $(\theta_1, \theta_{11})$  and a positive correlation between  $(\theta_1, \theta_{13})$  are observed. It can also be noted that no correlation is observed elsewhere, as assumed by the targeted density function (see Eq. (10)). Hence, the plots of  $\theta_{13}$  versus  $\theta_{11}$  draw orthogonal crosses (see scatterplot at row 3, column 2).

To conclude on this first exercise, we note that finding out the local optima is the most expensive stage. Identifying the multiple local optima is performed via a multi-start procedure launching searches from different initial locations in the parameter space. This procedure does not guarantee that all the local optima will be found. Our empirical experience shows however that the MCPD sampling (second stage) can identify some missed modes or improve the evaluation of some others. In this first stage, it is hard to state whether an optimization technique is a mark above the others. The use of gradient-based methods or evolutionary algorithms such as genetic algorithms or shuffled complex evolution methods are possible choices ([21]). For the second stage dealing with MCPDs sampling, the starting points in the parameter space are not far from the optimal locations. The calculations are

fast and optimizations relying on gradient-based methods are recommended because they converge rapidly when started close to the solution. In the present example, the computation time was also strongly reduced by launching independent MCPDs samplings using 25 processors. This resulted in a cost of 260 computational time units (CTU) for the step identifying the local optima, whereas MCPDs sampling took only 120 CTU.

## 5. Inverting a stochastic field with MCPDs

We consider here an inverse problem with a two-dimensional random field as the unknown parameters. The two-dimensional random field of scalar values is denoted by  $Y(\mathbf{x}, \omega)$  with  $\mathbf{x}$  the location in the Euclidean bounded space  $D$  and  $\omega$  a realization index of the random field equivalent to a coordinate in the probability space  $\Omega$ . Let us also consider a model response vector  $\mathbf{G}(Y(\mathbf{x}, \omega))$  and some observation data  $\mathbf{d}$ . The inverse problem consists of finding an optimal estimate  $\hat{Y}^{MAP}$  of the random field and its uncertainty range given the data. At first glance, the problem is ill-posed because there are an infinite number of unknowns, or at least, as many unknowns as the number of grid cells discretizing the domain  $D$  for numerically evaluating  $\mathbf{G}(Y)$ . Hence, the first task is to reduce the dimensionality (regarding parameters) of the problem.

### 5.1. Karhunen-Loève expansion

Several authors have suggested using Karhunen-Loève (KL) transformation of the random field ([13, 22–24]) to reduce the dimensionality of the problem. It is assumed that the random field obeys a second-order stationary Gaussian process, making that  $Y(\mathbf{x}, \omega) \sim \mathcal{GP}(\mu_Y, C_Y(\mathbf{x}_1, \mathbf{x}_2))$ , where  $\mu_Y$  is the mean of the process,  $(\mathbf{x}_1, \mathbf{x}_2)$  is a pair of different locations in the Euclidean space  $D$ , and  $C_Y(\mathbf{x}_1, \mathbf{x}_2)$  is the two-point covariance of the random



process. This covariance is a scalar continuous and positive-definite function corresponding to  $C_Y(\mathbf{x}_1, \mathbf{x}_2) = E[(Y(\mathbf{x}_1) - \mu_Y)(Y(\mathbf{x}_2) - \mu_Y)]$ , with  $E[\cdot]$  the mathematical expectation. Provided that  $Y(\mathbf{x}, \omega)$  is a real-valued random field with finite second moments, its KL expansion is

$$Y(\mathbf{x}, \omega) = \mu_Y + \sum_{i=1}^{+\infty} \sqrt{\lambda_i} \xi_i(\omega) \varphi_i(\mathbf{x}) \quad (11)$$

where the doublets  $(\lambda_i, \varphi_i(\mathbf{x}))$ ,  $\forall i \in \mathbb{N}^*$  are a set of eigenvalues ( $\lambda_1 \geq \lambda_2 \geq \dots > 0$ ) and the associated eigenfunctions of the covariance kernel  $C_Y$  while  $\{\xi_i\}_{i=1}^{\infty}$  are independent Gaussian random variables of zero mean and unit variance ( $\xi_i \sim \mathcal{N}(0, 1)$ ). The doublets eigenvalues-eigenfunctions are obtained by solving the Fredholm equation

$$\int_D C_Y(\mathbf{x}_1, \mathbf{x}_2) \varphi_i(\mathbf{x}_2) d\mathbf{x}_2 = \lambda_i \varphi_i(\mathbf{x}_1) \quad (12)$$

When ranked by decreasing values, the eigenvalues tend more or less rapidly to zero, thus allowing truncation of the KL development to the  $K^{\text{th}}$  order

$$\hat{Y}(\mathbf{x}, \omega) = \mu_Y + \sum_{i=1}^K \sqrt{\lambda_i} \xi_i(\omega) \varphi_i(\mathbf{x}) \quad (13)$$

where  $K$  is chosen so that  $\sum_{i=1}^K \lambda_i \geq (1 - \epsilon) \sum_{i=1}^{+\infty} \lambda_i$ . Given that the eigenfunctions  $\{\varphi_i(\mathbf{x})\}_{i=1}^{\infty}$  are continuous and form a complete orthonormal system in  $\mathcal{L}_2(D)$ , i.e.  $\int_D \varphi_i(\mathbf{x}) \varphi_j(\mathbf{x}) d\mathbf{x} = \delta_{ij}$  with  $\delta_{ij}$  the Kronecker delta function, Eq. (13) is therefore an expansion of  $Y(\mathbf{x}, \omega)$  onto an orthogonal basis. Provided the parameters  $\mu_Y$  and  $C_Y(\cdot)$  are fixed (guided by previous investigations), identifying  $\hat{Y}(\mathbf{x}, \omega)$  essentially involves seeking the plausible Gaussian random variables  $\xi^*$  that explain the observation data  $\mathbf{d}$ . The dimensionality of the inverse problem is therefore reduced to  $K$  parameters, i.e. the number of eigenvalues-eigenvectors chosen to expand  $Y(\mathbf{x}, \omega)$ . The problem is also partly regularized since  $\xi$  is a vector of independent Gaussian variables, i.e.

$p(\boldsymbol{\xi}) = \mathcal{N}(0, \mathbf{I}_K)$ . Note also that one can include the mean  $\boldsymbol{\mu}_Y$  in the sought parameters, as done in the sequel where one defines the parameter vector  $\boldsymbol{\theta} = (\boldsymbol{\mu}_Y, \boldsymbol{\xi})$  of dimension  $s = K + 1$ .

If we compare the above parameterization of a multiGaussian field with, for example a pilot point method, the reduction of dimensionality when solving the inverse problem is of the same order. Obtaining several plausible solutions in a Bayesian framework will finally require several optimizations procedures. On the one hand, a classical pilot point technique will duplicate the optimization of the pilot point values to modify different prior guesses on the parameter field. On the other hand, calculating MCPDs for the parameters of the Karhunen-Loève decomposition will also require multiple optimizations (see Section 3). Further studies handling carefully designed comparisons would be needed to identify advantages and drawbacks of both approaches.

### 5.2. Problem statement

Let us now illustrate the approach mixing a Karhunen-Loève decomposition and MCPD calculations. It must be first emphasized that the Karhunen-Loève decomposition is well known to depict accurately any second-order stationary random field ([23, 24]). Along this line, our study does not provide new insights on the KL decomposition. The main aim here is to assess the capability of MCPD samplings to correctly identify a large number of parameters that do not directly enter in the forward problem. As is the case with a lot of parameterization techniques, the optimized parameters serve as seeds to recompose the parameter field that enters in the calculation of the forward problem. In addition, the parameters of a Karhunen-Loève decomposition before inversion are independent because the eigenvectors of the decomposition are orthogonal. After inversion and post-conditioning, these

parameters may become correlated. It appears interesting to see how MCPD samplings will cope with this modification and which kind of uncertainty can be derived for parameters that are the seeds of a parameterization technique.

Let us now illustrate the approach by considering the following steady-state two-dimensional flow problem in a heterogeneous medium

$$\begin{cases} \nabla \cdot (T(\mathbf{x})\nabla h(\mathbf{x})) = 0, \mathbf{x} = (x, y) \in D = [0, 1] \times [0, 1] \\ \frac{\partial h(\mathbf{x})}{\partial y} \Big|_{y=0} = \frac{\partial h(\mathbf{x})}{\partial y} \Big|_{y=1} = 0 \\ T(\mathbf{x}) \frac{\partial h(\mathbf{x})}{\partial x} \Big|_{x=0} = 0.02; h(1, y) = 0 \end{cases} \quad (14)$$

where,  $h(\mathbf{x})$  [L] is the hydraulic head and  $T(\mathbf{x})$  [ $L^2T^{-1}$ ] is the transmissivity of the aquifer. The forward problem (14) is numerically solved by means of the mixed-hybrid finite element method ([25]) for a domain discretized into  $10^4$  triangular meshes.

The inverse problem consists of finding the log-transmissivity field  $Y(\mathbf{x}) = \log(T(\mathbf{x}))$  conditioned on a vector  $\mathbf{d}$  of data enclosing 25 measurements of hydraulic head spread all over the domain and 25 log-transmissivity values located at the same points as the observed heads. These data were obtained after running the flow scenario in Eq. (14) over a multiGaussian field  $Y(\mathbf{x})$  with a mean of zero and an isotropic Gaussian covariance  $C_y(\mathbf{x}_1, \mathbf{x}_2) = 2e^{-6|\mathbf{x}_1 - \mathbf{x}_2|^2}$ , yielding an effective correlation length of the random field  $Y(\mathbf{x})$  of 0.7. The data were then corrupted by a Gaussian white noise of standard deviation  $\sigma_h^m = 1$  for the heads and  $\sigma_Y^m = 0.1$  for the log-transmissivity. The reference field  $Y(\mathbf{x})$  and the measurement locations are reported in Fig. 4(a) and Fig. 4(d) respectively.

### 5.3. Posterior distribution

First, let us assume, as a prior guess, that the log-transmissivity field  $Y(\mathbf{x})$  has a mean of zero and an isotropic Exponential covariance  $C_y(\mathbf{x}_1, \mathbf{x}_2) =$

$10e^{-8|\mathbf{x}_1-\mathbf{x}_2|}$  yielding an effective correlation length of 0.375. This choice of an erroneous covariance function compared with that of the reference problem is twofold. First, we want to show that it is still possible to obtain a good estimation of the random field even though the prior guess on its spatial structure is flawed. For instance, the correlation length of the guess is here half the correlation length of the reference field, and the prior variance is five times the variance of the reference field. One can also mention to justify this discrepancy that the model of covariance in practical applications is often more conjectured than really known. Second, the Fredholm equation in (12) has well-known analytical solutions for an exponential covariance kernel (see [26]) which simplify and accelerate the calculations of the eigenvalues–eigenfunctions. It is of course possible to choose another covariance kernel. In that case, one can make use of the calculation method proposed in [27] to estimate the eigenmodes. In the present study, the KL expansion is performed by keeping the first 103 modes representing 84% of the variance associated with the field  $Y(\mathbf{x})$ . In the end, 104 parameters are sought (including the mean  $\mu_Y$ ), which is still a highly parameterized problem but far less than the initial problem that had a number of parameters equal to the number of meshes ( $10^4$ ) discretizing the domain.

Given the way the 50 local data were obtained, the likelihood function is written

$$p(\mathbf{d}|\boldsymbol{\theta}, \sigma_h^m, \sigma_Y^m) \propto \exp\left(-\frac{SS_h(\boldsymbol{\theta})}{2(\sigma_h^m)^2} - \frac{SS_Y(\boldsymbol{\theta})}{2(\sigma_Y^m)^2}\right) \quad (15)$$

where  $SS_h$  and  $SS_Y$  are the sum of squared errors on the hydraulic head and log-transmissivity respectively. In the present inversion exercise, we cannot include  $(\sigma_h, \sigma_Y)$  in the parameters to be estimated. The optimization algorithm overfits the data because the number of unknowns ( $s = 104$ ) is much greater than the number of data (i.e. 50). Hence, they are fixed to

their exact presumed value.

Keeping in mind that in the KL expansion,  $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathbf{I}_K)$  and assuming a uniform prior density function for  $\mu_Y$  allows to write (by using Eq. (15)) the parameter joint posterior density function as

$$p(\boldsymbol{\theta} | \mathbf{d}, \sigma_h^m, \sigma_Y^m) \propto \exp\left(-\frac{SS_h(\boldsymbol{\theta})}{2(\sigma_h^m)^2} - \frac{SS_Y(\boldsymbol{\theta})}{2(\sigma_Y^m)^2}\right) \prod_{i=1}^K \exp\left(-\frac{\xi_i^2}{2}\right) \quad (16)$$

Maximizing the above density function amounts to minimize the following weighted sum of squares,

$$J(\boldsymbol{\theta}) = \frac{SS_h(\boldsymbol{\theta})}{(\sigma_h^m)^2} + \frac{SS_Y(\boldsymbol{\theta})}{(\sigma_Y^m)^2} + \sum_{i=1}^K \xi_i^2 \quad (17)$$

The parameters  $\boldsymbol{\xi}$  being normally distributed a priori, we did not prescribe any restrictive variation ranges. Regarding the parameter  $\mu_Y$ , we prescribed a very large variation range  $[-100, 100]$  making in practice  $\mu_Y$  free of any constraint. The type of objective function in Eq. (17) is rapidly minimized with gradient-based methods if the Jacobian matrix  $\frac{\partial h(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  is calculated accurately ([28, 29]). A close look at the forward problem in (14) shows that its differentiation with respect to any parameter  $\boldsymbol{\theta}$  parameter yields

$$\left\{ \begin{array}{l} \nabla \cdot (T(\mathbf{x}, \boldsymbol{\theta}) \nabla h_\theta(\mathbf{x}, \boldsymbol{\theta})) = -\nabla \cdot (T_\theta(\mathbf{x}, \boldsymbol{\theta}) \nabla h(\mathbf{x}, \boldsymbol{\theta})), \mathbf{x} \in D \\ \frac{\partial h_\theta(\mathbf{x}, \boldsymbol{\theta})}{\partial y} \Big|_{y=0} = \frac{\partial h_\theta(\mathbf{x}, \boldsymbol{\theta})}{\partial y} \Big|_{y=1} = 0 \\ T(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial h_\theta(\mathbf{x}, \boldsymbol{\theta})}{\partial x} \Big|_{x=0} = -T_\theta(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial h(\mathbf{x}, \boldsymbol{\theta})}{\partial x} \Big|_{x=0}; h_\theta(1, y) = 0 \end{array} \right. \quad (18)$$

where the notations  $h_\theta(\mathbf{x}, \boldsymbol{\theta})$ ,  $T_\theta(\mathbf{x}, \boldsymbol{\theta})$  stand for the derivatives  $\frac{\partial h(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ ,  $\frac{\partial T(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ , respectively. Provided the forward problem has been solved and one can analytically differentiate the terms  $T_\theta(\mathbf{x}, \boldsymbol{\theta})$ , the calculation of (18) is very similar to that of the forward problem (14). It can be handled with the same code, the same discretization, resulting in comparable accuracy for evaluating both  $h(\mathbf{x}, \boldsymbol{\theta})$  and  $h_\theta(\mathbf{x}, \boldsymbol{\theta})$ .

#### 5.4. Results and discussion

With 104 coefficients kept in the KL expansion, MCPDs were calculated on an eight-core computer, each core calculating 13 MCPDs. On average, each core ran the model (forward flow problem + calculation of the Jacobian matrix) about 6 500 times (CTU) for a total over the eight cores of 54 560 runs. 105 spatially distributed problems were solved, one being the calculation of the head variable (Eq. (14)), and 104 evaluating sensitivities to parameters  $\theta_{-i}$  for a prescribed value of  $\theta_i$  (Eq. (18)). As a comparison, in [13], the authors handled a similar problem by means of MCMC and following the approach proposed in [12]. To lower the computational load, a surrogate model was used in a two-stage approach in which a new candidate must first pass the surrogate likelihood successfully before undergoing the statistical test of the original model. Three dependent chains were launched in parallel. The chains started to converge after 4 000 calls of the forward (i.e. original) model for a total of 10 000 calls (CTU).

The reference log-transmissivity field is reported in Fig. 4(a). The log-transmissivity field stemming from the KL expansion using the maximum a posteriori estimate of the parameter vector  $\boldsymbol{\theta}^{MAP} = (\mu_Y, \boldsymbol{\xi})^{MAP}$  is reported in Fig. 4(c). Both fields in 4(a) and 4(c) closely resemble each other, despite the fact that the covariance kernel of the KL expansion differs from that of the reference field (see above). Note also that we deliberately chose an erroneous covariance kernel for the KL expansion. However, the sampled covariances of the reference field and of the MAP estimates are very similar, implying that initial errors on the type of covariance and on the correlation length can be amended by inverting the KL coefficients and mixing eigenfunctions to recompose a relevant random field.

Fig. 4(b) also reports on the kriged map of the 25 local transmissivity

values. Though the patches of high and low values are correctly located, the kriged map differs from the reference field (Fig. 4(a)) and the MAP solution (Fig. 4(c)). For example, the effective correlation length seems overestimated and the overall map is smoother than the reference field. As shown later, the sampled variogram of the kriged map differs from that of MCPD solutions and confirms the preceding visual appraisals. Stated differently, the local transmissivity values do not conceal all the information and it makes sense inverting a flow problem conditioned on head data to retrieve the log-transmissivity field. It can also be noted that the inverse problems solved to calculate the MCPDs overfit the data, the number of parameters being larger than the conditioning data. The consequence is that the terms enclosed in the objective function Eq. (17) become lower than the prior measurement errors on heads and transmissivity data.

The uncertainty on the predicted log-transmissivity (or hydraulic head) can be estimated with Eq. (9). The uncertainty (the 95% confidence interval size) associated with the log-transmissivity field is shown in Fig. 4(d). The crosses mark the (co)locations of heads and transmissivities data used for the inversion. We can note that the uncertainty is not constant over the domain and is lower at the measurement locations where the 95% confidence interval size is about 0.4 (i.e.  $\sim \pm 1.96\sigma_Y^m$ ). This makes sense because the model prediction was anticipated to be more accurate at the measurement locations. Moreover, most of the domain is assigned an uncertainty range less than 0.8 ( $\Delta \log(T) < 0.8$ ) which is quite narrow except in the areas that contain no data (see North and South-West boundaries of the domain).

It is interesting to check whether the parameter uncertainty inferred via MCPDs yields a set of log-transmissivity fields including the reference (true) field. Fig. 4(e) maps the areas (in black) where the local value  $Y(\mathbf{x})$  of the

reference field is outside the 95% probable uncertainty range calculated by MCPDs (the width of these local uncertainty ranges are given in Fig. 4(d)). As expected, the areas where the uncertainty ranges do not include the local values of the reference field are those where no prior information was collected to solve the inverse problem. In the present case however, the uncertainty ranges are relatively narrow (Fig. 4(e)) and the non-matching areas represent 30% of the total domain. They reach 13% for the 99% predicted uncertainty range. Interestingly, the uncertainty bounds encompass the true field in the vicinity of observed data where the uncertainty range is smaller ( $\Delta \log(T) < 0.8$ ). This is the consequence of a well-conditioned problem. Observation data are rather evenly spread over the whole domain and very few sub-areas are not documented by at least a transmissivity value giving the order of magnitude of the parameters. Finally, it is worth noting that increasing the variance of the postulated Exponential covariance kernel (e.g.  $\sigma_Y^2 = 20$ ) provides slightly wider uncertainty bounds that finally encompass the true field (not shown). This artificial increase of the  $\sigma_Y^2$  value to obtain increased uncertainties on the transmissivity field is not a consequence of flawed MCPD samplings. The handled problem is here highly regularized by the Karhunen-Loève decomposition which controls the freedom of action onto the tuned parameters (i.e. the eigenvalues of the decomposition).

The MCPD estimates of the first KL coefficients are reported in Fig. 5 ("diagonal" plots). We notice that the density of  $\mu_Y$  has a bell-shaped curve centered on zero, though a prior uniform density was postulated. Fig. 5 also reports, in its "off-diagonal" plots, on the various draws of parameters  $\boldsymbol{\theta}$  to build MCPDs. A plot in column  $i$  and row  $j$  corresponds to the optimal values of the parameter  $\theta_j$  for sampled (prescribed) values of  $\theta_i$  and optimal values of  $\theta_i$  for sampled  $\theta_j$ . These plots for the pair  $(\mu_Y, \xi_1)$  yield a unique straight



line of non-null slope indicating a strong negative correlation between these two parameters. Notably,  $\mu_Y$  and  $\xi_1$  are also closely correlated with  $\xi_6$  (row #7, columns #1-2 of Fig. 5) meaning that these parameters have dependent effects on the state variable calculated by the flow model. We remind that the prior guess on the parameters  $\xi$  assumed that they were independent and of Gaussian distribution with zero mean and unit variance. The posterior estimates show that some parameters can be highly correlated and that their distribution may have changed. For example, the parameters from  $\xi_2$  to  $\xi_4$  show bell-shaped MCPDs, but are clearly non-centered on zero and have a variance less than one (width of the bell-shaped curve far less than 6).

The overall quality of the inverse solutions is now discussed. First, the MAP solution reported in Fig. 4(c) has a spatial structure close to that of the reference solution. Note that the reference solution has an effective correlation length of about 0.5 which is smaller than expected from the covariance model used to generate the random field. This occurs when the size of the domain is close to the correlation length. The sampled covariance (variogram) of the MAP solution is Gaussian with a correlation length of about 0.45, close to that of the reference field (see Fig. 6(a)). The MAP solution, however, slightly underestimates the variance, probably because the KL expansion considers modes representing only 84% of the variance on  $Y$  (see above). However, the MAP solution is much closer from the reference field than a kriged map of transmissivity data. This kriged map has a variogram that largely overestimates the correlation length and underestimates the variance of the transmissivity field. As told above the kriged field is smoother than the reference and MAP solutions. We note that some inverse solutions calculated by taking parameters from MCPDs sampling can perfectly fit the covariance of the reference field (Fig. 6(a)). These solutions are good but not

the best with respect to the posterior pdf.

Note that MCPDs sample give optimal inverse solutions for a prescribed value of one parameter. Using the sample, we rebuilt the associated  $Y$  fields and calculated their sampled (experimental) covariances. Then, these covariances were fitted with a Gaussian model in the form of  $C_y(\mathbf{x}_1, \mathbf{x}_2) = \sigma_y^2 e^{-\eta(\mathbf{x}_1 - \mathbf{x}_2)^2}$ , the variance of  $C_Y$  obviously being  $\sigma_y^2$  and the effective correlation length  $\sqrt{3/\eta}$ . Fig. 6(b), (c), and (d) show these variances and correlation lengths. In general, the MCPD solutions slightly underestimate the variance of  $Y$  with mean values on the order of 1.9 for a reference at 2.1. The mean correlation length of MCPD solutions establishes at 0.43 for a reference at 0.47. Though the prior guess on the spatial structure of the inverse solutions greatly differed from the reference, the eigenvectors using the KL expansion clearly record the effective spatial structure (covariance) from which data are extracted. The re-composition of these eigenvectors by means of MCPDs provides a large set of probable solutions that are only very slightly biased, at least for the forward problem handled here.

## 6. Conclusions

In this work, we introduced the concept of maximal conditional posterior distribution for the identification of model parameters. The MCPD of a parameter  $\theta_i$  can be viewed as the distribution of the parameter  $\theta_i$  knowing that the other parameters are at their optimal values (in the sense that the conditional posterior pdf is maximized). It makes sense calculating a MCPD because the uncertainty drawn from the distribution of  $\theta_i$  refers to valuable solutions to the inverse problem. All the parameters  $\theta_{-i}$  are optimal except  $\theta_i$  which is allowed to wander in the parameter space. In essence, the calculations of two MCPDs for parameters  $\theta_i$  and  $\theta_j$  are independent.

They can be performed very easily on parallel computational sessions and one can extend the number of parallel sessions up to the total number of parameters involved in the problem under investigation (thus, drastically reducing the computational effort). This feature associated with the fact that the MCPD draws are good inverse solutions post-conditioned on observation data, renders the inversion in a Bayesian context affordable, even for highly parameterized problems.

As stated above, a MCPD allows a single parameter  $\theta_i$  to vary when the others are optimal. A key feature to avoid the single parameter (and the others) to be far from valuable solutions is to pre-identify the so called maximum a posteriori (MAP). This MAP is a solution where all the parameters are at their optimal value, but this solution may correspond to several local maxima when the underlying distribution of parameters is multimodal. Hence, the success of MCPD sampling depends on the ability of the preliminary optimization procedure to retrieve all the probable local optima. This can be achieved by multi-start optimizations. The synthetic test case performed in this study showed that the MCPD calculations retrieve fairly well multimodal distributions.

The MCPD technique was also faced to the problem of retrieving a parameter random field on the basis of information gathering both measurements of the parameters and state variables from a spatially distributed problem (here, steady-state groundwater flow). The parameterization allowing a strong decrease in the dimensionality (in parameters) of the problem was grounded in a Karhunen-Loève decomposition of the parameter field. This technique is well-known to be accurate regarding the depiction of a stationary random field. Hence, the MCPD calculations were really tested on their ability to provide valuable inverse solutions and the associated uncertainties for pa-

rameters that do not directly enter in the calculation of the forward problem (the parameters allow the re-composition of a transmissivity field then used in the forward groundwater flow problem). The whole procedure allowed us to approximate the random field with only  $s = 104$  parameters: the mean of the random field and the 103 first eigenmodes of the Karhunen-Loève decomposition. Then, the MCPDs of these parameters were assessed quickly with the proposed algorithm taking advantage of their computation in parallel.

It was shown that the MCPD-KL association was able to accurately retrieve a reference field even when starting from a flawed initial estimate of the spatial structure (covariance) of the field. MCPDs can also render parameter distributions that strongly differ from their prior guess especially regarding their mean and variance.

### **Acknowledgement**

The authors are grateful to the French National Research Agency who funded this work (project RESAIN n°ANR-12-BS06-0010-02).

## References

- [1] N.-A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equations of state calculations by fast computing machines, *Journal of Chemical Physics* 21 (1953) 1087–1091. [doi:10.1063/1.1699114](https://doi.org/10.1063/1.1699114).
- [2] H. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* 57 (1970) 97–109. [doi:10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97).
- [3] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, J. Tu, A framework for validation of computer models, *Technometrics* 49 (2007) 138–154. [doi:10.1198/004017007000000092](https://doi.org/10.1198/004017007000000092).
- [4] D. Higdon, J. Gattiker, B. Williams, M. Rightley, Computer model calibration using high-dimensional output, *J. American Statistical Association* 103 (2008) 570–583. [doi:10.1198/016214507000000888](https://doi.org/10.1198/016214507000000888).
- [5] H. Haario, Saksman, J. Tamminen, An adaptive metropolis algorithm, *Bernoulli* 7 (2001) 223–242. [doi:10.2307/3318737](https://doi.org/10.2307/3318737).
- [6] P. J. Green, A. Mira, Delayed rejection in reversible jump metropolis-hastings, *Biometrika* 88 (2001) 1035–1053. [doi:10.1093/biomet/88.4.1035](https://doi.org/10.1093/biomet/88.4.1035).
- [7] H. Haario, M. Laine, A. Mira, E. Saksman, Dram: Efficient adaptive mcmc, *Statistics and Computing* 16 (339–354). [doi:10.1007/s11222-006-9438-0](https://doi.org/10.1007/s11222-006-9438-0).
- [8] C. J. F. ter Braak, J. Vrugt, Differential evolution markov chain with

- snooker updater and fewer chains, *Statistics and Computing* 18 (4) (2008) 435–446. [doi:10.1007/s11222-008-9104-9](https://doi.org/10.1007/s11222-008-9104-9).
- [9] M. C. Kennedy, A. O’Hagan, Bayesian calibration of computer models, *Journal of the Royal Statistical Society* 63 (B) (2001) 425–464. [doi:10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294).
- [10] J. A. Christen, C. Fox, Mcmc using an approximation, *Journal of Computational and Graphical Statistics* 14 (4) (2005) 795–810. [doi:10.1198/106186005X76983](https://doi.org/10.1198/106186005X76983).
- [11] X. Ma, N. Zabaras, An efficient bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, *Inverse Problems* 25 (2009) 035013. [doi:10.1088/0266-5611/25/3/035013](https://doi.org/10.1088/0266-5611/25/3/035013).
- [12] T. Cui, C. Fox, M. A. O’Sullivan, Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance metropolis-hastings algorithm, *Water Resources Research* 47 (2011) W10521. [doi:10.1029/2010WR010352](https://doi.org/10.1029/2010WR010352).
- [13] E. Laloy, B. Rogiers, J. A. Vrugt, D. Mallants, D. Jacques, Efficient posterior exploration of a high-dimensional groundwater model from two-stage markov chain monte carlo simulation and polynomial chaos expansion, *Water Resources Research* 49 (2013) 2664–2682. [doi:10.1002/wrcr.20226](https://doi.org/10.1002/wrcr.20226).
- [14] M. Loève, *Probability Theory*, Springer-Verlag, Berlin, 1977.
- [15] A. Tarantola, *Inverse Problem. Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.

- [16] C. J. F. ter Braak, A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces, *Statistics and Computing* 16 (2006) 239–249. [doi:10.1007/s11222-006-8769-1](https://doi.org/10.1007/s11222-006-8769-1).
- [17] J. A. Vrugt, C. J. F. ter Braak, C. G. H. Diks, D. Higdon, B. A. Robinson, J. M. Hyman, Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *International Journal of Nonlinear Sciences and Numerical Simulation* 10 (2009) 273–290. [doi:10.1515/IJNSNS.2009.10.3.273](https://doi.org/10.1515/IJNSNS.2009.10.3.273).
- [18] J. A. Vrugt, C. J. F. ter Braak, H. V. Gupta, B. A. Robinson, Equifinality of formal (dream) and informal (glue) bayesian approaches in hydrologic modeling?, *Stochastic Environmental Research and Risk Assessment* 23 (7) (2008) 1011–1026. [doi:10.1007/s00477-008-0274-y](https://doi.org/10.1007/s00477-008-0274-y).
- [19] M. Laine, Adaptative mcmc methods with applications in environmental and geophysical models, Ph.D. thesis, Finnish Meteorological Institute, Finland (2008).
- [20] E. Laloy, J. A. Vrugt, High-dimensional posterior exploration of hydrologic models using multiple-try dream<sub>(zs)</sub> and high-performance computing, *Water Resources Research* 48 (2012) W01526. [doi:10.1029/2011WR010608](https://doi.org/10.1029/2011WR010608).
- [21] Q. Duan, S. Sorooshian, V. K. Gupta, Optimal use of the sce-ua global optimization method for calibrating watershed models, *Journal of Hydrology* 158 (1994) 265–284. [doi:10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4).
- [22] W. Li, A. Cirpka, Efficient geostatistical inverse methods for structured

- and unstructured grids, *Water Resources Research* 42 (2006) W06402. [doi:10.1029/2005WR004668](https://doi.org/10.1029/2005WR004668).
- [23] Y. Efendiev, T. Y. Hou, W. Luo, Preconditioning markov chain monte carlo simulations using coarse-scale models, *SIAM Journal on Scientific Computing* 28 (2006) 776–803. [doi:10.1137/050628568](https://doi.org/10.1137/050628568).
- [24] Y. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems, *Journal of Computational Physics* 228 (6) (2009) 1862–1902. [doi:10.1016/j.jcp.2008.11.024](https://doi.org/10.1016/j.jcp.2008.11.024).
- [25] A. Younes, P. Ackerer, R. Mose, G. Chavent, A new formulation of the mixed-finite element method for solving elliptic and parabolic pde with triangular elements, *Journal of Computational Physics* 149 (1999) 148–167. [doi:10.1006/jcph.1998.6150](https://doi.org/10.1006/jcph.1998.6150).
- [26] D. Zhang, Z. Lu, An efficient high-order perturbation approach for flow in random porous media via karhunen–loève and polynomial expansions, *Journal of Computational Physics* 194 (2004) 773–794. [doi:10.1016/j.jcp.2003.09.015](https://doi.org/10.1016/j.jcp.2003.09.015).
- [27] K. K. Phoon, S. P. Huang, S. T. Quek, Implementation of karhunen-loeve expansion for simulation using a wavelet-galerkin scheme, *Probabilistic Engineering Mechanics* 17 (2002) 293–303. [doi:10.1016/S0266-8920\(02\)00013-9](https://doi.org/10.1016/S0266-8920(02)00013-9).
- [28] A. Kaczmaryk, F. Delay, Improving dual-porosity-medium approaches to account for karstic flow in a fractured limestone. application to the automatic inversion of hydraulic interference tests, *Journal of Hydrology* 347 (2007) 349–403. [doi:10.1016/j.jhydrol.2007.09.027](https://doi.org/10.1016/j.jhydrol.2007.09.027).



- [29] P. Ackerer, F. Delay, Inversion of a set of well-test interferences in a fractured limestone aquifer by using an automatic multi-scale parameterization technique, *Journal of Hydrology* 389 (2010) 42–56. [doi:10.1016/j.jhydrol.2010.05.020](https://doi.org/10.1016/j.jhydrol.2010.05.020).

## Figure captions

### List of Figures

- 1 A two-dimensional illustration of the MCPD assessment. The MCPD of  $\theta_1$  is built by maximizing the conditional joint posterior distribution. During this maximization, the most probable value of  $\theta_2$  is derived while the value of  $\theta_1$  is fixed. This operation is repeated by fixing  $\theta_1$  farther and farther from its maximum a posteriori estimate . . . . . 35
- 2 General algorithm used to infer MCPDs in the vicinity of the MAP, assuming that no parallelization of the calculations is performed (3.2 for explanations). Specifically,  $N_{it}$  corresponds to the maximal number of iterations used to refine the sampling of a MCPD (part below the green lozenge in the middle of the chart);  $d$  is the number of parameters. Notably, when the algorithm seeks MCPDs for multimodal densities, the sequence in the chart is simply resumed for each local  $\theta^{opt}$  (just replace  $\theta^{MAP}$  by  $\theta^{opt}$ ). . . . . 36
- 3 The posterior parameter densities of a multivariate Gaussian distribution are represented on the diagonal. The blue solid lines are the MCPDs while the marginal posterior densities are in red broken lines. The MCPD of  $\theta_{13}$  is very different from its marginal density because two modes overlap. The off-diagonal plots represent some of the MCPD draws. For a row  $i$  and a column  $j$ , horizontal direction corresponds to the values of parameter  $\theta_i$  and vertical direction to the values of  $\theta_j$ . Plots of optimal  $\theta_j$  values for MCPD draws of  $\theta_i$  and plots of optimal  $\theta_i$  for draws of  $\theta_j$  . . . . . 37

4	<p>The true log-transmissivity field are represented in (a). The kriged estimated field is represented in (b). In (c) the MAP estimate of the log-transmissivity field is depicted. The size of the 95% uncertainty interval of the estimated field is shown in (d), the crosses indicate the measurement locations. In (e), the black stains indicate the region of the domain where the estimated uncertainty log-transmissivity field does not encompass the true field . . . . .</p>	38
5	<p>The estimated MCPDs of the first seven coefficients in the KL approximation of the log-transmissivity field (on-diagonal plots). The off-diagonal plots represent some of the MCPD draws. For a row <math>i</math> and a column <math>j</math>, horizontal direction corresponds to the values of parameter <math>\theta_i</math> and vertical direction to the values of <math>\theta_j</math>. Plots of optimal <math>\theta_j</math> values for MCPD draws of <math>\theta_i</math> and plots of optimal <math>\theta_i</math> for draws of <math>\theta_j</math> . . . . .</p>	39
6	<p>Comparison of <math>Y = LogK</math> variograms between the reference problem and MCDP inversions. (a)- variograms of the reference field, the MAP field and one of the MCPDs fields. (b), (c), variances and correlation lengths of variograms from all MCPD fields with respect to MCPD probability values. The red line is the location of the reference field. (d)- correlation length versus variance of variograms from MCPD fields. The red dot corresponds to the reference field . . . . .</p>	40

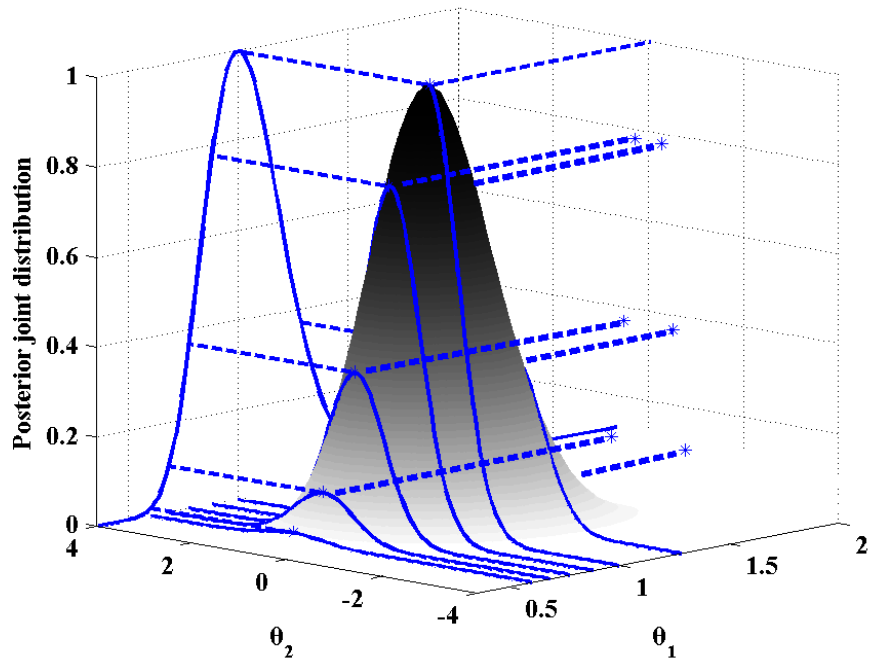


Figure 1:

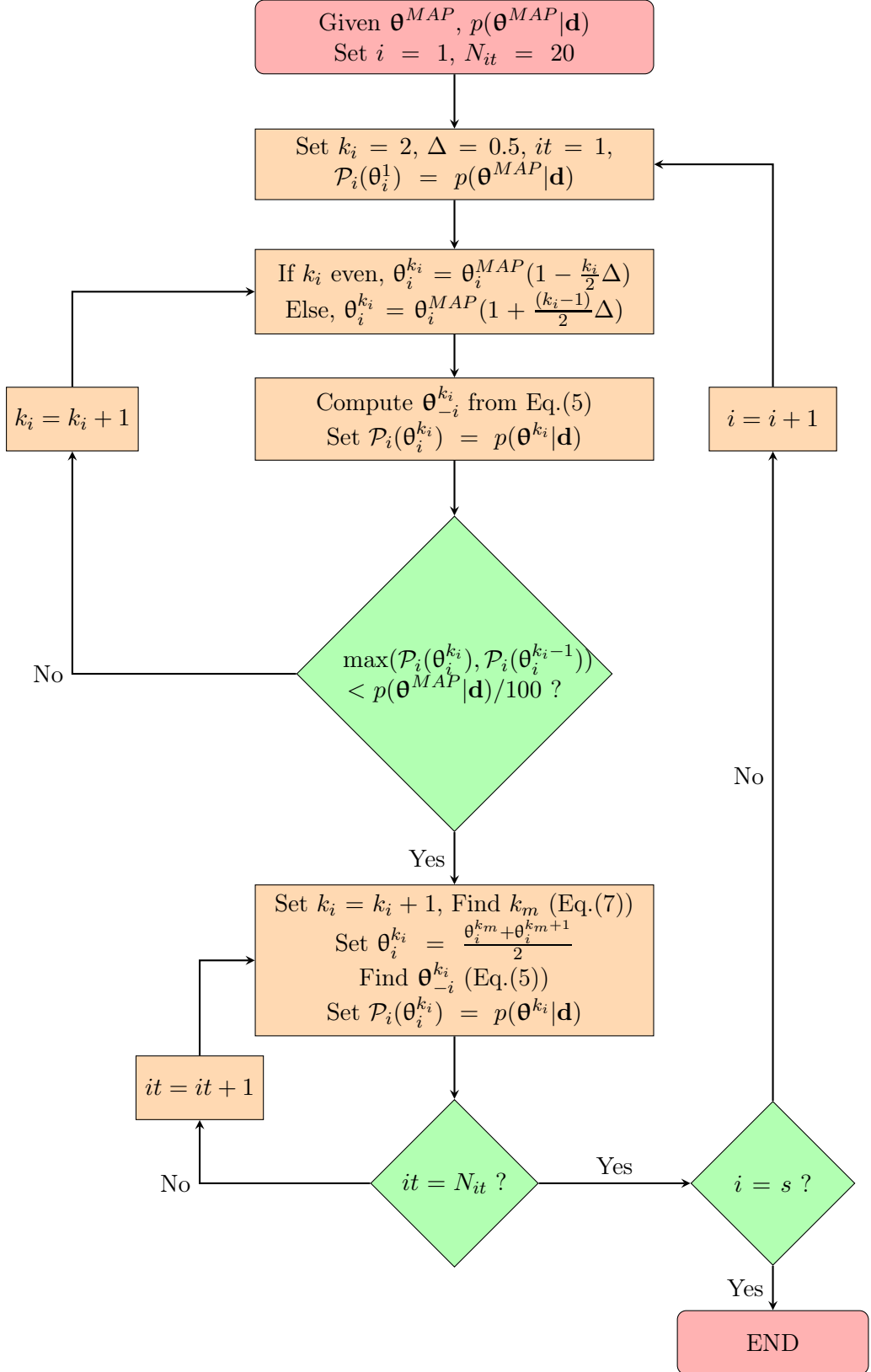


Figure 2:

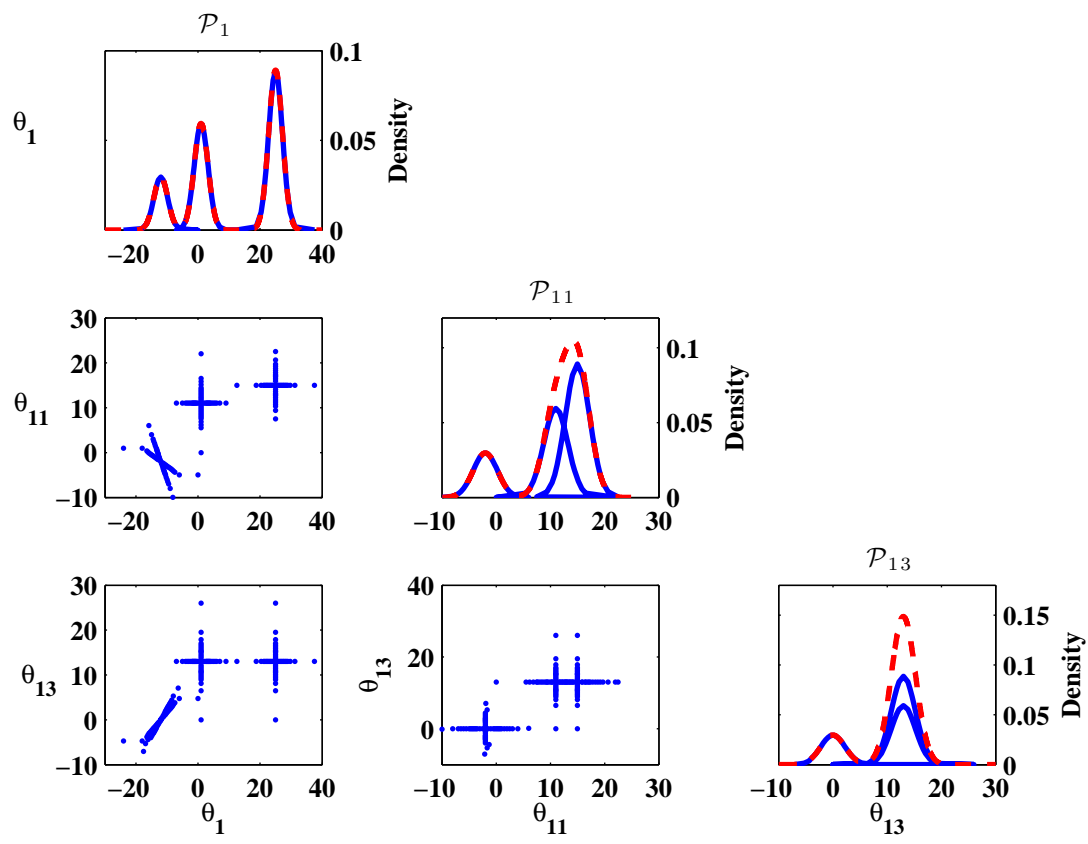


Figure 3:

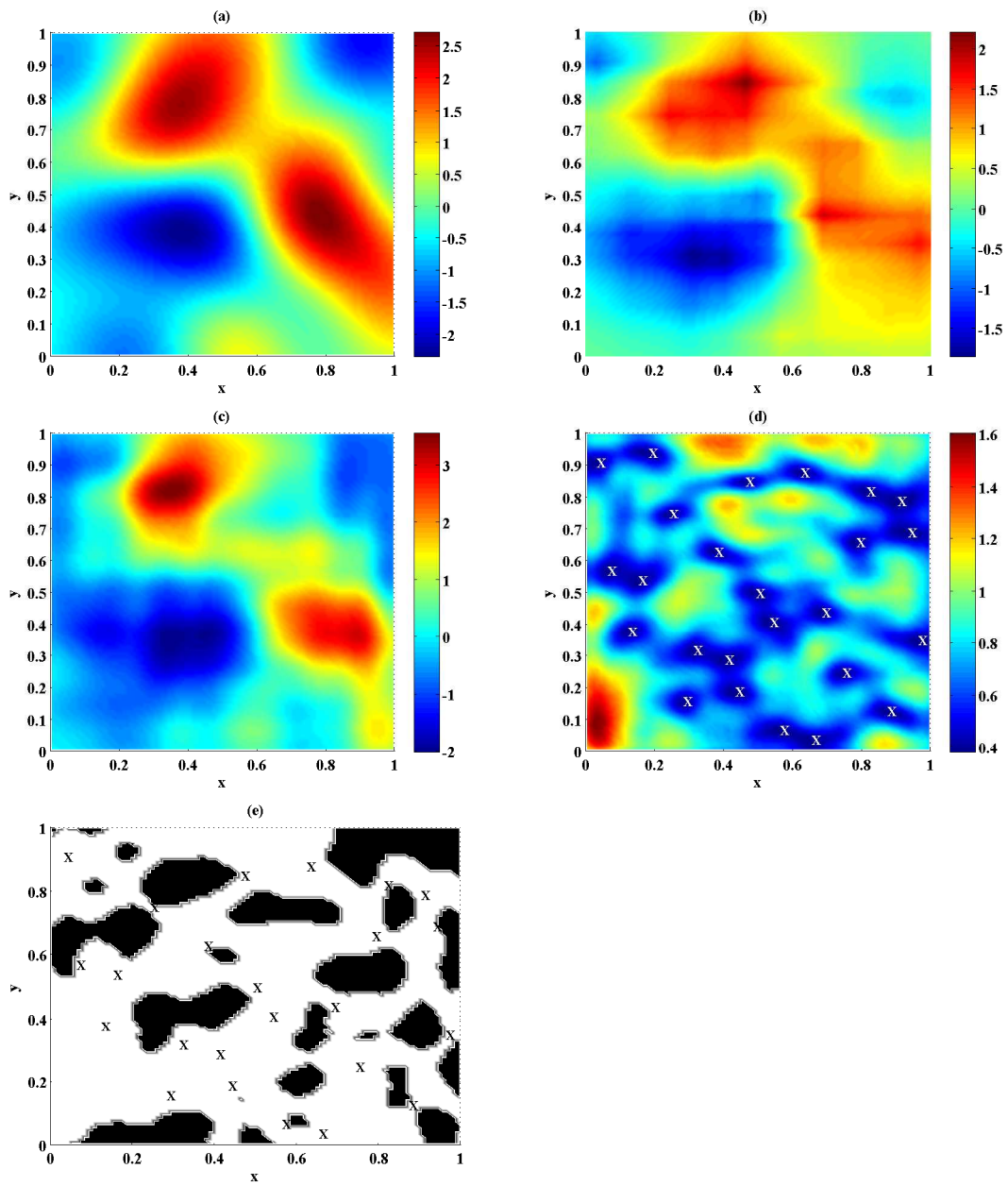


Figure 4:

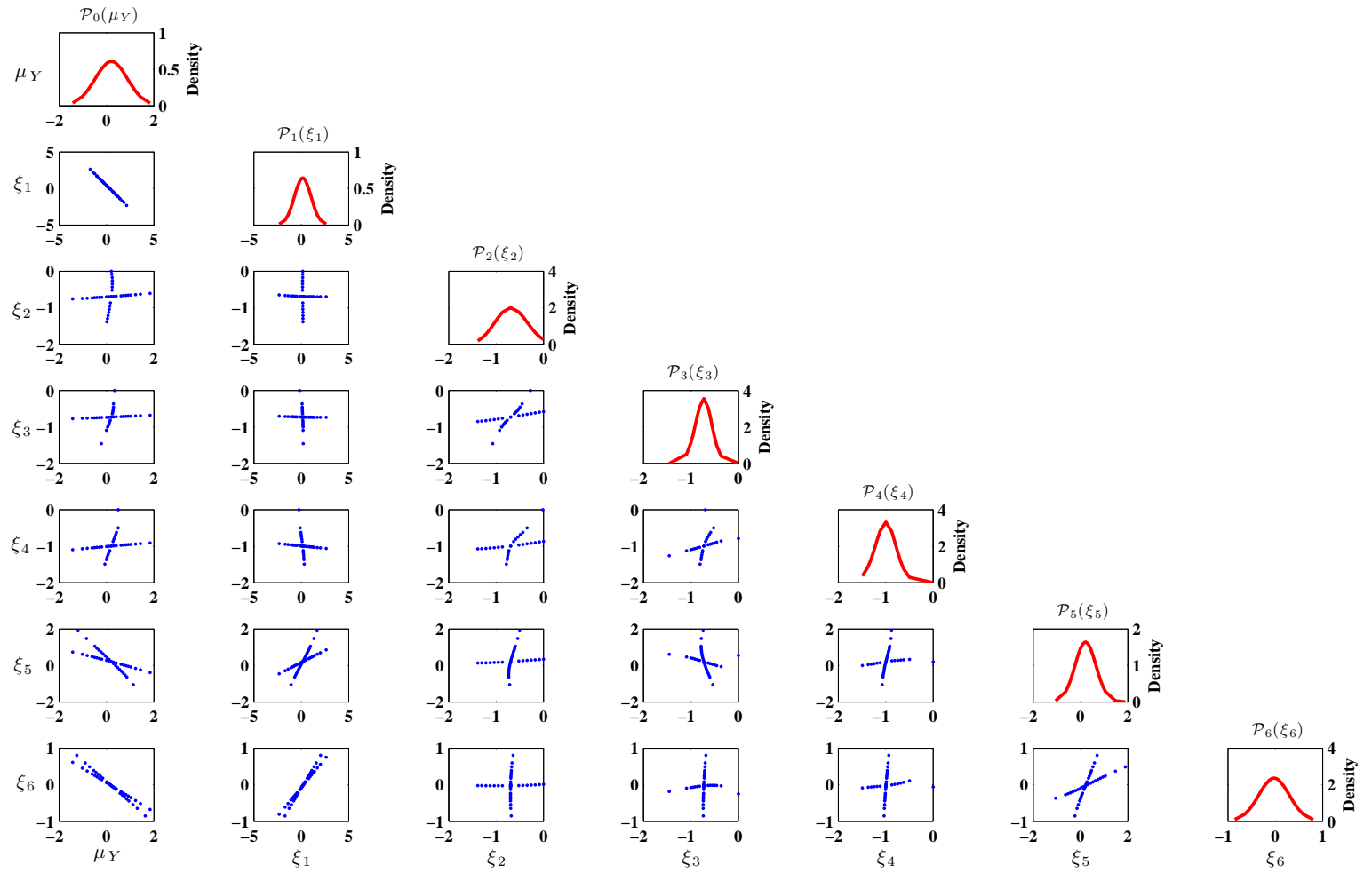


Figure 5:



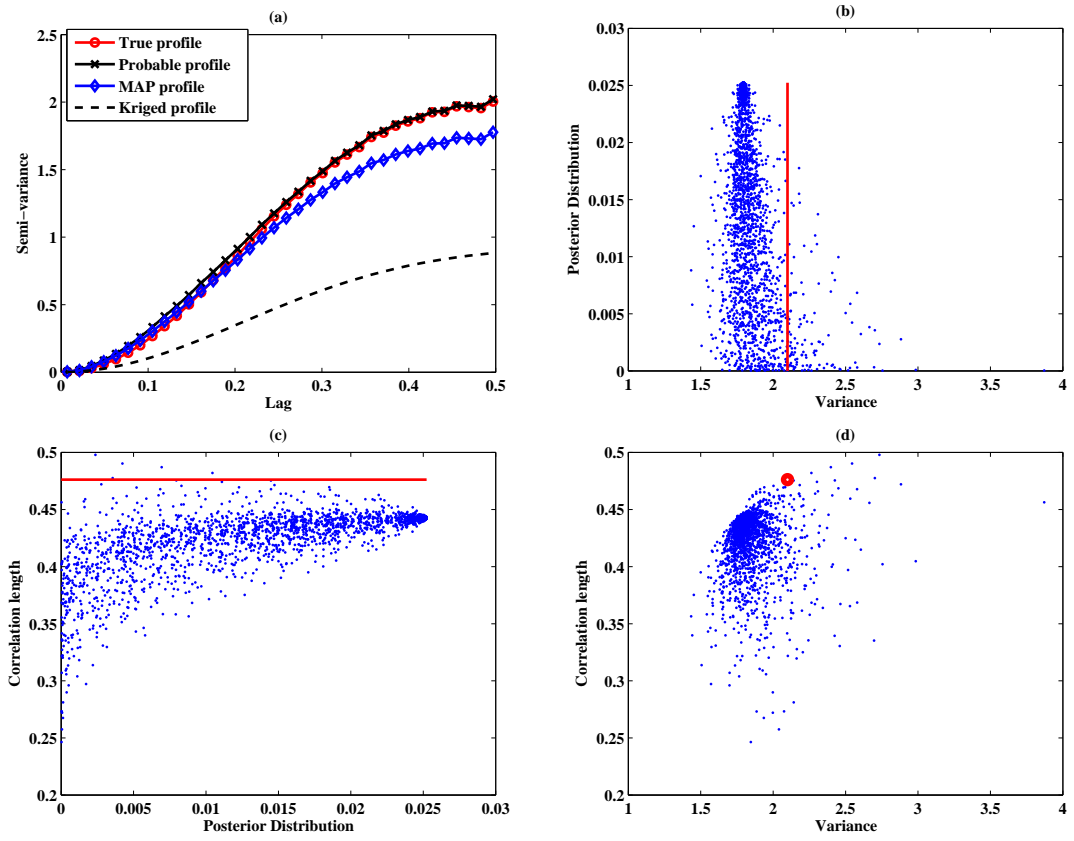


Figure 6: