Comput Geosci (2012) 16:1–20
DOI 10.1007/s10596-011-9249-z

REVIEW PAPER

A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow

Ángel Yustres · Laura Asensio · Juan Alonso · Vicente Navarro

Received: 2 September 2010 / Accepted: 12 July 2011 / Published online: 3 August 2011

Abstract Parameter identification is one of the key elements in the construction of models in geosciences. However, inherent difficulties such as the instability of ill-posed problems or the presence of multiple local optima may impede the execution of this task. Regularization methods and Bayesian formulations, such as the maximum a posteriori estimation approach, have been used to overcome those complications. Nevertheless, in some instances, a more in-depth analysis of the inverse problem is advisable before obtaining estimates of the optimal parameters. The Markov Chain Monte Carlo (MCMC) methods used in Bayesian inference have been applied in the last 10 years in several fields of geosciences such as hydrology, geophysics or reservoir engineering. In the present paper, a compilation of basic tools for inference and a case study illustrating the practical application of them are given. Firstly, an introduction to the Bayesian approach to the inverse problem is provided together with the most common sampling algorithms with MCMC chains. Secondly, a series of estimators for quantities of interest, such as the marginal densities or the normalization constant of the posterior distribution of the parameters, are reviewed. Those reduce the computational cost significantly, using only the time needed to obtain a sample of the posterior probability density function. The use of the information theory principles for the experimental design and for the ill-posedness diagnosis is also introduced. Finally, a case study based on a highly instrumented well test found in the literature is presented. The results obtained are compared with the ones computed by the maximum likelihood estimation approach.

Keywords Inverse problem · Markov Chain Monte Carlo · Information theory · Uncertainty quantification

Mathematics Subject Classifications 49N45 · 65C05 · 62F15 · 62B10 · 62G07 · 62P12 · 62P30

Á. Yustres (✉) · L. Asensio · J. Alonso · V. Navarro
Geoenvironmental Group, Civil Engineering School, University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: angel.yustres@uclm.es,

L. Asensio
e-mail: laura.asensio@uclm.es

J. Alonso
e-mail: juan.alonso@uclm.es

V. Navarro
e-mail: vicente.navarro@uclm.es

1 Introduction

In any branch of natural science, any attempt of modelization requires determining the mathematical formulation of the model and estimating its parameters. This process is generally known as inverse modeling [1]. Solving this problem is especially important in the field of geosciences, since the high degree of heterogeneity of the physical environment requires a characterization of the spatial distribution of the parameter values. Hence only the measurements of the variable to be modeled will be available ([1], chapter 12.4).

In most cases, the model structure is determined a priori, in keeping with the previous knowledge of the modeler. Thus, the estimation of the parameters is the only thing to be done to define a single model. However, the selection of over-parameterized models may result in poor prediction capability, although they

may allow measured data to be reproduced without bias ([2], chapter 1.4). The trade-off between the number of parameters and the prediction capability may be solved through the maximum likelihood estimation method [3] and information theory. Akaike [2, 4, 5] proposed a criterion for the classification of models up for consideration in terms of AIC criterion. In its simplest expression, it requires the value of the likelihood function evaluated at its maximum. Some other criteria, such as the BIC [6], the Minimum Description Length [7], the KIC [8] or the YIC [9, 10] have been extensively used in hydrological modeling [11–14]. In addition, another family of model selection techniques based exclusively on Bayesian Theory, such as the Bayes factors, has been proposed [15–17]. This methodology has also been applied in hydrological sciences [18–21], together with model averaging approach for multimodel inference [22, 23]. However, model selection and averaging is beyond the scope of this paper, and we will concentrate here on parameter estimation as the only component of inverse modeling.

Despite the many possible ways of considering the inverse problem, most of them have a similar mathematical formulation based on the construction and analysis of a merit function [24–26]. Among all the possible analyses, the optimization of that function is the most usual alternative. In this approach, the parameters of a model are usually considered to be fixed and unique. Nevertheless, in more advanced applications, the parameters can be considered as time varying [27–32]. Along with the error structure, the parameters define a single probability density function (PDF) of which the measured data are occurrences. Each of these realizations has an associated PDF value, which is a function of the parameters to be determined. The combination of those PDFs, which is called likelihood [33], is given by

$$L(\mathbf{d}|\mathbf{m}) = \prod_{i=1}^{n} \lambda(d_i|\mathbf{m}) \tag{1}$$

where $\mathbf{d}$ is the vector of observations (or data vector), d_i is the ith observation, $n$is the number of observations, $\mathbf{m}$ is the vector of the model parameters (or parameter vector) and λ is the PDF of each observation conditioned on the selected model $\mathbf{d} = \mathbf{g}(\mathbf{m})$. An estimator of the value of the parameters with a lot of nice statistical properties, such as consistency and asymptotic normality and efficiency [34, 35], is the maximum likelihood estimator (MLE), the point at which the likelihood function exhibits its maximum. If the errors have a Gaussian structure, the MLE estimator becomes the point that minimizes the sum of squared errors weighted by their variance–covariance matrix. This gives rise to a geometric interpretation of the maximum likelihood principle, which can be interpreted as a minimization of the weighted distance between the observed and the simulated values.

This approach should be used only if the problem does not present non-identifiable parameters, as long as the necessary and sufficient conditions [36, 37] are met. The non-identifiability of a set of parameters is defined [38, 39] as the non-dependence of the likelihood function $L(\mathbf{d}|\mathbf{m})$ on them. Usually the data vector does not provide any information on these parameters. In such cases, Tikhonov regularization is advised [1, 25, 40], although solutions are expected to be biased [41].

Another perspective of the inverse problem relies on the Bayesian probability theory. According to this theory, the axioms of probability can be interpreted in terms of a subjective degree of knowledge [42–44]. The uncertainty associated with a parameter or a particular field measurement can be expressed in the form of a PDF. From this point of view, the inverse problem becomes a combination of our states of knowledge on the parameters [45–47], including prior information. The solution to the inverse problem is not based on obtaining an estimator of the parameter vector. Rather, it provides an objective state of belief which is represented by a PDF.

In the field of mathematics, both approaches have evolved independently owing to theoretical controversies. However, the use of prior information in geosciences is quite common [25, 26, 48], although it is based on the point estimation of the parameters [49] by the Maximum a posteriori (MAP) approach [50, 51], rather than on Bayesian inference. In fact, the use of Gaussian prior information is a regularization procedure similar to the one proposed by Tikhonov ([26, 52], chapter 4.3; [53]). It provides solutions with little computational cost, especially in cases where a large amount of variables are involved (heterogeneity at grid cell level). However, the analysis of the results obtained is carried out using local approximations of the posterior covariance matrix around the optimal value of the posterior PDF (MAP approach) [25, 46], which may be an unrepresentative point in multimodal distributions [54]. These PDFs are relatively frequent in environmental sciences modeling [55], but may present as well non-smoothness. This feature may be caused by, among other factors, the time stepping scheme used for solving the differential equations involved in the problem [56]. Consequently the MAP approach would be useful only in well-behaved functions [57].

This paper reviews both conceptual and numerical tools to provide a complete analysis of the inverse

problem from a Bayesian perspective, with special focus on the Markov Chain Monte Carlo methods. These methods have been used for 20 years in several fields of geosciences, although they are not frequent in everyday practice. This work reviews the Bayesian theory of inverse problems, as well as a series of existing algorithms that will allow for a detailed analysis of the posterior PDF of the model parameters. Furthermore, the reviewed algorithms reduce the computational expense incurred in the calculation of marginal densities and integrals.

2 Bayesian approach of inverse problems

As mentioned in the introduction, the Bayesian approach to inverse problems is based on the interpretation of probability as states of information [42–44], where all knowledge is expressed in the form of probability density functions. Any kind of plausible reasoning can only be carried out by the combination of these states of information [43, 45].

Only two sources of information are available to resolve an inverse problem: (1) information that provides an analysis of the observed values (data) on the basis of a theoretical relation (model) and (2) our previous knowledge of the model parameters and data. [58]. The Bayesian formulation of the posterior PDF of the parameters is given by the expression [51, 58, 59]

$$p\left(\mathbf{m}|\,\mathbf{d}\right) = \frac{p\left(\mathbf{d}|\,\mathbf{m}\right) p\left(\mathbf{m}\right)}{\int_{\mathfrak{M}} p\left(\mathbf{d}|\,\mathbf{m}\right) p\left(\mathbf{m}\right)} \tag{2}$$

where $p(\mathbf{d}|\mathbf{m})$ is the likelihood function (equivalent to the expression in Eq. 1), $p(\mathbf{m})$ is the prior probability of the parameters and the integral in the denominator is the normalization constant, also referred in the literature as $p(\mathbf{d})$.

If $p(\mathbf{m})$ and the observation errors are multivariate normal, then $p(\mathbf{m}|\mathbf{d})$ may be expressed as [46, 51]:

$$p\left(\mathbf{m}|\,\mathbf{d}\right) = k \cdot \exp\left(-\frac{1}{2}\left(\left(\mathbf{g}(\mathbf{m}) - \mathbf{d}\right)^T \mathbf{C}_\mathrm{D}^{-1}\left(\mathbf{g}(\mathbf{m}) - \mathbf{d}\right) + \left(\mathbf{m} - \mathbf{m}_\mathrm{o}\right)^T \mathbf{C}_\mathrm{M}^{-1}\left(\mathbf{m} - \mathbf{m}_\mathrm{o}\right)\right)\right) \tag{3}$$

where k is the normalization constant, $\mathbf{d}$ is the vector of observed values, $\mathbf{C}_\mathrm{D}$ is the covariance matrix of the errors, $\mathbf{m}_\mathrm{o}$ is the mean of $p(\mathbf{m})$, and lastly, $\mathbf{C}_\mathrm{M}$ is the covariance matrix of $p(\mathbf{m})$. In the event that $\mathbf{C}_\mathrm{M}^{-1} = \mathbf{0}$ (non-informative prior information), the point at which the $p(\mathbf{m}|\mathbf{d})$ is maximized will be the vector of parameters that minimizes the weighted sum of squares of the first term of Eq. 3. When a parameter (or a set of them) is non-identifiable [38, 39] the use of prior information is indispensable. However, if the priors are also non-informative for the non-identifiable parameters, the problem is considered ill-posed [38]. Furthermore, priors that not integrate to unity or contain little information on the parameters of interest may also induce ill-posedness as defined by Hadamard [38, 60].

In accordance with Eq. 3, the pure maximum likelihood estimation ignores the existence of any prior information. In the past, this triggered some controversy between the frequentist and Bayesian theories of probability [42, 61]. To a degree, this controversy has been resolved in geosciences by using regularization approaches that include, to a certain extent, prior information [1, 25, 26].

If the variance–covariance matrices in Eq. 3 are not known a priori, their components can be considered as unknown parameters which have their own prior densities. The same occurs in the case of the direct model $\mathbf{d} = \mathbf{g}(\mathbf{m})$, where the measurement of the forcing data (precipitation or temperature in the case hydrological modeling for example) is subjected to error. For a complete analysis of all the sources of uncertainty, the hierarchical modeling approach proposed by the BATEA framework is advised, especially if a complete differentiation between input, response and model error is desired [30, 62, 63].

Equation 3 provides the modeler with a PDF that synthesizes all the information that can be drawn from the observed values and from our prior knowledge [43, 64]. By means of integration, it is possible to obtain from $p(\mathbf{m}|\mathbf{d})$, the moments of any order, marginal PDFs of any parameter or set of parameters, as well as their graphical representations in low dimensional spaces.

3 Markov Chain Monte Carlo sampling methods

Despite the fact that a PDF is available when following the Bayesian approach, in geosciences it is customary to analyze problems that involve a large number of parameters. Consequently, the classical methods of integration and representation may be too costly in computational terms. If we also consider that, in most cases, $p(\mathbf{m}|\mathbf{d})$ lacks a simple analytical expression of their descriptors, Monte Carlo methods are advisable for use in this type of analysis [48].

The term Monte Carlo method stands for any member of a very large group of algorithms used to solve various kinds of computational problems by means of randomly generated numbers [65, 66]. Of the many existing methods to obtain a sample of $p(\mathbf{m}|\mathbf{d})$, perhaps the most appropriate would be the MCMC method.

The basic principle of these techniques is the application of the convergence property to a specific probability density function presented by the Markov Chains, provided that they comply with the conditions of reversibility and ergodicity [67]. This means that if it is possible to develop an algorithm that generates an ergodic reversible Markov Chain over a continuous space, then the states (points) of this chain will converge asymptotically to a sample of the function considered, also called stationary distribution.

The first algorithm that was applied to the generation of these Markov Chains was the so-called Metropolis–Hastings Sampler [67–69]. This sampling method obtained a new state of the Markov Chain (point in the space to be sampled) in two steps: a proposal step and an acceptance step. An example of a posterior density function like $p(\mathbf{m}|\mathbf{d})$ would be

3.1 Proposal step

- Start out with the state of the Markov Chain $X_i = \mathbf{m}_i$
- A sample point $\mathbf{m}_c$ is generated from the proposal PDF $q(\mathbf{m}_i, \mathbf{m}_c)$ which defines the probability that the Markov chain is found in $\mathbf{m}_i$; go to point $\mathbf{m}_c$.

3.2 Acceptance step

- Calculate the ratio $\alpha\,(\mathbf{m}_i, \mathbf{m}_c) = \min\left(1, \frac{p(\mathbf{m}_c|\mathbf{d})q(\mathbf{m}_c,\mathbf{m}_i)}{p(\mathbf{m}_i|\mathbf{d})q(\mathbf{m}_i,\mathbf{m}_c)}\right)$
- If $\alpha\,(\mathbf{m}_i, \mathbf{m}_c) = 1$, then $\mathbf{m}_c$ is accepted as the new state of the Markov Chain $X_{i+1} = \mathbf{m}_{i+1} = \mathbf{m}_c$
- If $\alpha\,(\mathbf{m}_i, \mathbf{m}_c) < 1$, then a random number, ρ, is obtained from a uniform distribution over [0, 1). If $\rho < \alpha\,(\mathbf{m}_i, \mathbf{m}_c)$ then $\mathbf{m}_c$ is accepted as the new state of the Markov Chain $X_{i+1} = \mathbf{m}_{i+1} = \mathbf{m}_c$. If the opposite occurs, then the state of the Markov Chain does not change.
- Begin a new proposal step, either with a new initial state X_{i+1} (if there is acceptance) or with the original one (X_i in the event of rejection)

As can be observed, the simplicity of this algorithm makes it especially useful for the sampling of posterior probability density functions. Moreover the Metropolis–Hastings sampler presents two other fundamental advantages. The first is that if $q(\mathbf{m}_i, \mathbf{m}_c) = q(\mathbf{m}_c, \mathbf{m}_i)$, it will hold that

$$\alpha\,(\mathbf{m}_i, \mathbf{m}_c) = \min\left(1, \frac{p\,(\mathbf{m}_c\,|\mathbf{d}\,)}{p\,(\mathbf{m}_i\,|\mathbf{d}\,)}\right). \tag{4}$$

Therefore, to calculate $\alpha(\mathbf{m}_i, \mathbf{m}_c)$ it is only necessary to evaluate the PDF at points $\mathbf{m}_i$ and $\mathbf{m}_c$. The other advantage is that it is not even necessary to know the normalization constant k, but only the ratio of the posterior density at $\mathbf{m}_i$ and $\mathbf{m}_c$. That means, if $p(\mathbf{m}|\mathbf{d}) = kp^*(\mathbf{m}|\mathbf{d})$, the value of α is obtained as follows

$$\alpha\,(\mathbf{m}_i, \mathbf{m}_c) = \min\left(1, \frac{p^*\,(\mathbf{m}_c\,|\mathbf{d}\,)}{p^*\,(\mathbf{m}_i\,|\mathbf{d}\,)}\right) \tag{5}$$

which allows for the sampling of PDFs even without being normalized.

Due to their versatility, the MCMC methods have been used in hydrological sciences and geophysics [70–77]. Although the Metropolis–Hastings sampler is very well suited for most of those applications (only the unnormalized value of the posterior PDF is needed), some other sampling strategies have been proposed. In the case of Gibbs sampler [78, 79] it is necessary to know any conditional probability density of the original PDF, in order to be sampled. This requirement is easy to fulfill in closed form expressions for the prior densities, but it can be very computationally expensive for the likelihood term. This fact is particularly important for the numerical distributed models widely applied in geosciences [80] where no closed form is available for the posterior conditionals. However, the use of genetic operators as proposal functions that has given rise to the Evolutionary MCMC algorithms [81, 82] is probably the most promising computational tool for the sampling unnormalized PDFs. The differential evolution adaptive metropolis (DREAM) Markov Chain Monte Carlo (MCMC) scheme [83, 84] has been applied to hydrological sciences due to its better convergence and mixing properties [85]. The randomized maximum likelihood algorithm [51, 76, 86] also presents a better rate of acceptance of the Metropolis–Hastings criterion shown in Eq. 5, although it requires the resolution of multiple optimization problems.

As was previously stated, Markov Chains converge asymptotically to their stationary distribution. Several methods have been developed to evaluate the degree of convergence, ranging from graphical methods [79] to more complex statistical methods [87–89]. However, despite the solid foundations of all of these methods, it has been demonstrated that any method when applied individually may lead to poor performance. Hence it is advisable to use a combination of them all [90].

To achieve a low rejection rate (points rejected versus points evaluated) of the algorithm and a rapid convergence to the posterior PDF, several different sampling strategies may be used [91]. However, perhaps the most effective technique is problem reparametrization to reduce the dimensionality of the problem, or at least to reduce the parameter correlation [92]. This paper has selected a proposal function $q(\mathbf{m}_i, \mathbf{m}_c)$ that bears the greatest resemblance to a $p(\mathbf{m}|\mathbf{d})$, which allows for better mixing and faster convergence. In order to do this,

it is necessary to take a small initial sample, so that the process is restarted with $q(\mathbf{m}_i, \mathbf{m}_c) \sim N\left(\hat{\mu}_s, \hat{\mathbf{C}}_s\right)$, where $\hat{\mu}_s$ and $\hat{\mathbf{C}}_s$ are the mean and covariance matrix of the initial sample respectively. The process is repeated until an acceptable degree of convergence [88] is reached.

Another issue in the implementation of MCMC samplers is the computational burden required, especially in problems with time-dependent parameters such as those arisen in environmental engineering. The use of numerical solutions for the nonlinear differential equations that describe the physical process makes the parameter vector strongly dependent upon previous values. If the BATEA [62] or other hierarchical Bayesian approach is applied, multiple evaluations of the posterior PDF are needed and the computational cost would increase prohibitively. A "limited memory" MCMC sampler has been proposed to overcome this issue [80]. This is done by truncating the temporal dependence when the "memory effect" of the previous conditions falls below a threshold value.

The inverse problem in subsurface flow presents an additional difficulty: the parameterization of the geological heterogeneity [51]. The MCMC sampling procedure may become inefficient when a large number of parameters (heterogeneity at grid cell level) is being considered. The blocking MCMC [75] or the use of coarse scale models [77, 93, 94] are based on the use of upscalling techniques in order to reduce the number or parameters.

Further attention should be devoted to dynamical systems, where new data are being continuously obtained. The process of incorporating new information as it becomes available is known in literature as data assimilation. The Kalman filter and the extended Kalman filter (for nonlinear problems) are the most accepted methods for this assimilation [51]. However, the extension of ensemble Kalman filtering [51] with MCMC methods [95] presents very promising results for data assimilation and uncertainty quantification in subsurface problems.

4 Inference from Markov Chain Monte Carlo samples

4.1 Marginal density estimation

Once a sample of the posterior density has been obtained, inference from the MCMC output can be drawn. Since the sample usually shows an autocorrelation with a variable lag depending on $q(\mathbf{m}_i, \mathbf{m}_c)$ it is not advisable to use the classical estimators of the mean and variance of the PDF. Some alternative estimators have been proposed ([67, 88, 96], chapter 12.4). Nevertheless, frequently it does not suffice to rely only on the intuitive idea provided by the estimators or the scatter plots. In such cases it would be advisable to estimate the marginal probability density at any point on the basis of the MCMC sample obtained. Some non-parametric marginal density estimators have been proposed [91]—namely, the conditional marginal density estimator [97]. This estimator requires full knowledge of any conditional probability of $p(\mathbf{m}|\mathbf{d})$ up to its normalization constant. This is a major drawback, which can be avoided by using the importance weighted marginal density estimator (IWMDE) [98], since this estimator only requires the conditional probability densities of an arbitrary weight function. This function was chosen with two factors in mind. The first is that its form must resemble $p(\mathbf{m}|\mathbf{d})$ as closely as possible. The second requirement is that its conditional probability densities must be easy to calculate. The IWMDE at point $\mathbf{m}^{*(j)}$ is obtained by the expression:

$$p_{\mathrm{IWMDE}}\left(\mathbf{m}^{*(j)}|\mathbf{d}\right) = \frac{1}{N}\sum_{i=1}^{N} w\left(\mathbf{m}_i^{(j)}\middle|\mathbf{m}_i^{(-j)}\right)\frac{p\left(\mathbf{m}^{*(j)}, \mathbf{m}_i^{(-j)}|\mathbf{d}\right)}{p\left(\mathbf{m}_i|\mathbf{d}\right)} \tag{6}$$

where $p_{\mathrm{IWMDE}}\left(\mathbf{m}^{*(j)}|\mathbf{d}\right)$ is the proposed estimator, N is the total number of points of the MCMC sample, $w(\mathbf{m})$ is the weight function, $\mathbf{m}_i$ is the ith point of the sample and $p\left(\mathbf{m}^{*(j)}, \mathbf{m}_i^{(-j)}|\mathbf{d}\right)$ is the value of $p(\mathbf{m}|\mathbf{d})$ at the point where the marginal PDF has been estimated $\left(\mathbf{m}^{*(j)}\right)$, and where the rest of the coordinates are those of the point $\mathbf{m}_i^{(-j)}$. Thus, index (j) denotes the group of variables in which the IWMDE is obtained and index $(-j)$ indicates the rest of the components of vector $\mathbf{m}$. Therefore the value of $p(\mathbf{m}|\mathbf{d})$ up to its normalization constant is not needed since it is cancelled out in ratio $p(\mathbf{m}^{*(j)},\mathbf{m}_i^{(j)}|\mathbf{d})/p(\mathbf{m}_i|\mathbf{d})$.

The most outstanding of the non-parametric methods is the kernel density estimation (KDE) [99–101]. It is based on computing a sum equal to the length of the sample of bounded density functions on the space, called kernels. The kernel estimator at point $\mathbf{m}^{*(j)}$ may be calculated as

$$p_{\mathrm{K}}\left(\mathbf{m}^{*(j)}|\mathbf{d}\right) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{\mathbf{m}^{*(j)} - \mathbf{m}_i^{(j)}}{h}\right) \tag{7}$$

where N is the sample size, h is a scale factor and $K(\mathbf{m}^{(j)})$ is the kernel function. Since kernel density estimators are non-parametric, they may not be efficient. However, with an adequate choice of the kernel function, good results may be obtained at a very low cost

[102]. The resulting computational efficiency may be attributed to the fact that, unlike the IWMDE, no additional evaluation of the PDF is required. This makes it especially well-suited for being used with the MCMC output, even if it is not possible to evaluate the posterior PDF again.

4.2 Normalization constant

The value of the normalization constant k is not needed to obtain a sample of the posterior density $p(\mathbf{m}|\mathbf{d})$ or its marginal PDFs. Nonetheless, it may be required for the computation of Bayes Factors and for the multimodel inference proposed in the Bayesian Model Averaging methods [16, 17, 103, 104]. For this reason, while it is possible to use the classic Monte Carlo integration methods [105] or importance sampling [106], there are algorithms that use only the MCMC output (posterior sample and values of the PDF at those points) [107]. The normalization constant can be estimated as

$$\log k = \log L\left(\mathbf{m}^*\right) - \log\left[\frac{1}{M}\sum_{i=1}^{M}\frac{\omega\left(\mathbf{m}_i\right)}{p\left(\mathbf{m}_i\right)}\frac{L\left(\mathbf{m}^*\right)}{L\left(\mathbf{m}_i\right)}\right] \tag{8}$$

where $L(\mathbf{m}^*)$ is the value of the likelihood function at an arbitrary point $\mathbf{m}^*$ (preferably a high posterior density point), $[\mathbf{m}_i]$ is the MCMC sample and M corresponds to its size. The weight function $\omega(\mathbf{m})$ must be a probability density function which consequently will ensure that $\int_{\mathfrak{M}} \omega\left(\mathbf{m}\right) d\mathbf{m} = 1$. Since the evaluation of $\omega(\mathbf{m})_i$ is generally much less computationally costly than $L(\mathbf{d}|\mathbf{m}_i)$ and $\mathrm{p}(\mathbf{m}_i)$, the time required to obtain $\log k$ is negligible in comparison with the time needed to obtain the posterior sample.

4.3 Posterior diagnostics

Once the sample of the posterior PDF has been obtained, a critical examination of the assumptions made during the formulation of the inverse problem is required [108, 109]. Firstly, an assessment on the compatibility between the prior and the posterior information is desirable. If both densities are too far apart, probably either the prior information is not reliable or the likelihood function is not well constructed. A scatter diagram of the posterior and prior sample can help in this process, as well as for detecting non-identifiable parameters or strong correlations between them.

Another aspect that should be assessed is the theoretical distributions of the errors in the model. In many cases the normality of the residuals is a strong hypothesis that should be examined. For that purpose, it is advised to use a normal probability plot of the residuals [110] and non-parametric tests such as the Kolmogorov–Smirnov test.

Finally, the well-posedness of the problem should be assessed. Usually this feature cannot be uncovered by formal mathematical methods [38] in the case of complex models. For that reason, indirect empirical evidences of ill-posedness are sought, such as the condition number of the Hessian matrix at the MLE [55, 63], the Fisher information matrix [111, 112] or the poor convergence of the MCMC samplers [38, 87]. The following section reviews the Shannon information theory and its use in posterior ill-posedness diagnostics.

5 Information theory in inverse problem analysis

The Monte Carlo Methods and the Bayesian approach to inverse problems allow us to use global measures of the information. That is, the complete posterior density is taken into account, and not only the value of the likelihood function at the MLE and the posterior covariance matrix. The total amount of information contained in a PDF, such as for example, $p(\mathbf{m}|\mathbf{d})$, is defined as [113]:

$$I_{\mathrm{S}} = \int_{\mathfrak{M}} p\left(\mathbf{m}\,|\mathbf{d}\right)\log\left(\frac{p\left(\mathbf{m}\,|\mathbf{d}\right)}{\mu\left(\mathbf{m}\right)}\right)d\mathbf{m} \tag{9}$$

where $\mu(\mathbf{m})$ is the uniform PDF. With this expression, called Shannon's information measure (which is the inverse of the Shannon entropy), it is possible to evaluate the information using $\mu(\mathbf{m})$ as a reference. The units in which the information is expressed are called bits when the base-2 logarithm is used, and nats or nits when a natural logarithm is used. For $\mu(\mathbf{m})$ it is true that $I_{\mathrm{S}} = 0$, which in Bayesianstatistics is called non-informative. However, it is possible to understand the expression proposed by Shannon as a particular case of the expression [114]:

$$I_{\mathrm{KL}} = \int_{\mathfrak{M}} p\left(\mathbf{m}\,|\mathbf{d}\right)\log\left(\frac{p\left(\mathbf{m}\,|\mathbf{d}\right)}{\phi\left(\mathbf{m}\right)}\right)d\mathbf{m} \tag{10}$$

which may be interpreted intuitively as the distance that exists between the probability density functions $p(\mathbf{m}|\mathbf{d})$ and $\phi(\mathbf{m})$. This measure is non-symmetric. Hence, it is called Kullback-Leibler (KL) divergence, rather than distance, and it assesses the dissimilarity between two density functions [114]. Therefore, I_S may be interpreted as a measurement of the dissimilarity between $p(\mathbf{m}|\mathbf{d})$ and $\mu(\mathbf{m})$, where $\mu(\mathbf{m})$ designates an absolute reference that contains null information. The KL divergence underlies most of the model selection criteria mentioned in the introduction [2, 5].

The calculation of Shannon's information measure is based on the integration of the posterior density in the entire domain, as opposed to Fisher's information, which is based on the analysis of the second derivatives around the optimum. As regards multimodal PDFs, the differences between the two information measures are relevant, since the former may be considered as a global measure and the latter, a local one. The integral of Eq. 9, may be calculated by means of importance sampling [105, 106], a classical method of Monte Carlo integration, also proposed for the calculation of normalization constants of the PDFs and consequently of their Bayes factors [115, 116]. The importance function used is $p(\mathbf{m}|\mathbf{d})$ itself, since the sample obtained from the MCMC sampler is available. Given that $p(\mathbf{m}|\mathbf{d})$ is overdispersed with respect to the integrand of Eq. 9, but it does have the same features as said integrand, the numerical integral converges quickly. To evaluate I_S it is necessary to know the posterior density up to its normalization constant, k. That is, $p^*(\mathbf{m}|\mathbf{d})$ must first be integrated using the algorithm proposed in Eq. 8. The normalization constant k and the information content I_s can be calculated without any additional evaluation of the posterior PDF, using only the MCMC output.

The information content of the posterior PDF can play a dual role in the analysis of inverse problems. Firstly, the marginal information of the parameters can be used as an indicator of ill-posedness of the inverse problem (or non-identifiability if no prior information is available [38]). If a parameter exhibits a low value of marginal information, in contrast to the rest of them, it is likely that the problem can be considered ill-posed. For unimodal PDFs this criteria is quite similar to the condition number of the Hessian matrix at the MAP estimator. However, this evaluation of the ill-posedness is not conclusive, and other methods of diagnosis have been proposed, such as the study of the correlation of the MCMC samples or the lack of convergence of the sampling algorithm [38]. The combination of all of them is probably the best alternative to the formal mathematical analysis [38] when this is not possible due to the complexity of the model. Moreover, the identification of the parameters with low marginal information in a non-identifiable problem can be helpful for the search of highly informative priors for them.

The second application of information theory is the quantitative evaluation of different experimental alternatives for data acquisition (design of monitoring networks, frequency of measurement or selection of variables of interest) based on the total and marginal information provided by each experimental design. This capability helps overcome the problem of non-identifiability by the selection of the most informative design. In addition, the quantification of the information provided by any (additional) data is also possible [117] as well as, for example, the determination of the duration of the observation period or the selection of observations to be analyzed.

6 A practical example

6.1 Description of the problem

To illustrate the sampling process, an analysis of a very conventional inverse problem, like the pumping test, is proposed. Although these kinds of tests may not provide reliable estimates of the average parameter values of an aquifer [118], they have been widely used in hydrogeology with satisfactory results. The case proposed here, an unconfined aquifer test conducted in a glacial outwash deposit on Cape Cod, Massachusetts, was studied. The geological formation under consideration was a homogeneous, anisotropic unconfined aquifer [119]. A partially penetrating well was drilled and was pumped at a rate of 2.02×10^{-2} m^3/s for 72 h. Up to 20 piezometers were used to monitor the water level during the test, apart from the pumped well itself. A complete description of the experimental setup can be found in the literature [119–121].

The case selected here is especially appropriate for analysis by means of the Monte Carlo methods and the information theory. First of all, the test was carried out in a highly homogeneous aquifer. Hence, the analytical solution selected may be used without violating

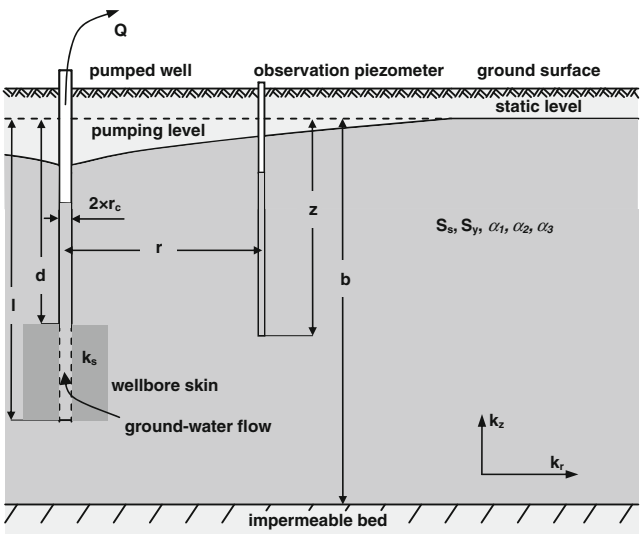

Fig. 1 Pumping test configuration and variables involved in Moench's analytical solution

Table 1 Limits of the search space

	Lower limit	Upper limit
k_r (m/s)	1.000×10^{-9}	1.000×10^{-1}
k_z (m/s)	1.000×10^{-9}	1.000×10^{-1}
k_s (m/s)	1.000×10^{-9}	1.000×10^{-1}
b (m)	3.346×10^{1}	1.000×10^{3}
S_y (-)	1.000×10^{-6}	1.000×10^{0}
S_s (m^{-1})	1.000×10^{-9}	1.000×10^{0}
α_1 (s^{-1})	1.000×10^{-10}	1.000×10^{10}
α_2 (s^{-1})	1.000×10^{-10}	1.000×10^{10}
α_3 (s^{-1})	1.000×10^{-10}	1.000×10^{10}

its initial assumptions. Moreover, the test was highly instrumented, including a large number of piezometers, and the test conditions were under control. Also, the test lasted long enough to provide early-time as well as intermediate and late time data. The aquifer has also been studied in depth and the test interpreted on the basis of the least-squares approach mentioned previously. With this method it is possible to compare the results obtained and to evaluate the degree of correctness found in the uncertainty estimation. Lastly, the complete report of Moench et al. [119] offers a detailed description of the experimental procedure, the analytical solution used in the interpretation and the different hypotheses put forward in the analysis of the available information.

Nine parameters have been proposed for the resolution of the inverse problem (see Fig. 1). Since the aquifer in question is anisotropic, both radial and vertical permeability, k_r and k_z (m/s), are two of the fundamental parameters. The hydraulic conductivity of the wellbore skin, k_s (m/s), must also be identified. In reality, k_s lacks significance if the thickness of the wellbore skin, d_s (m) is unknown. Therefore, the value to be identified would be the non-dimensional factor $S_w = k_r \cdot d_s / k_s \cdot r_w$, where r_w (m) is the outside radius of the pumped well screen. This paper considers $d_s = r_w$ so that k_s includes all of the information on the two parameters. Moreover, it is necessary to identify two fundamental parameters of the unconfined aquifer analyzed, i.e., the specific storage, S_s (m^{-1}), and the specific yield, S_y (–). Also taken into account are the parameters, α_1, α_2 and α_3 (s^{-1}), which define the delayed

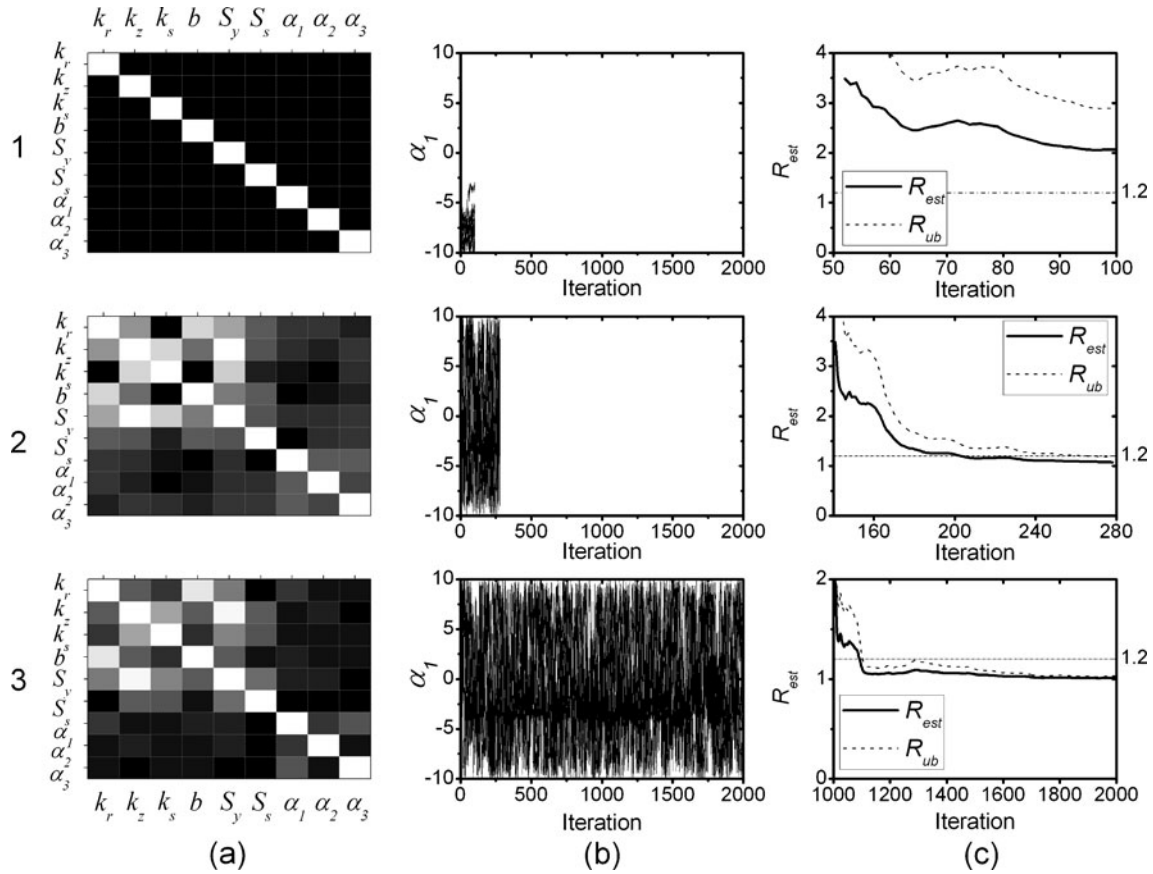

Fig. 2 Three stages sampling procedure. **a** Absolute values of the correlation matrix of the. proposal function (*White* = 1, *Black* = 0). **b** chain evolution for α_1. In the first stage, the chains got stuck in a reduced area of the parameter space. **c** Convergence diagnostics based on Gelman et al. [88]. The value of $R = 1.2$ denotes the start of convergence, an $R = 1.0$ the total convergence

drainage from the vadose zone in the analytical solution proposed [119, 122, 123]. Finally, the original saturated thickness, b (m), is included, even though there are previous estimates of roughly 49 m [124–126].

The resolution of the joint inverse problem [119] was carried out by means of nonlinear least squares software, containing a gradient-based optimization algorithm plus a series of tools for the local analysis of the uncertainty of the estimated parameters [127]. Since direct optimization might be complicated in a problem that is highly multidimensional, nonlinear and probably ill-posed, the process was carried out in several steps. Firstly, only late-time data and very few variables were considered. Subsequently, more variables were gradually introduced and the early-time data were used. In this way, a fairly adequate initial parameter vector was obtained to be taken as a starting point in the resolution of the joint problem. This led to the start of an optimization process which produced the numerical values of eight out of the nine parameters under consideration, as S_w was determined independently, while the other remained fixed.

6.2 Sampling strategy

As a preliminary approximation, samples of the posterior PDFs of each individual piezometer and the pumping well were obtained, together with the joint inverse problem (using the complete data set) [119] within the upper and lower limits shown in Table 1. In all of the cases, a non-informative prior with those limits was included in the Bayesian formulation and $\mathbf{C}_D = \hat{\sigma}_e^2\mathbf{I}$ was considered, where $\hat{\sigma}_e^2$ designates the estimated value of the error variance from Moench's estimated parameters and $\mathbf{I}$ is the identity matrix.

The Metropolis–Hastings algorithm was used to sample the posterior parameter distribution of each individual piezometer. For this purpose a multi-stage process was employed to tune the sampler. The approach proposed by Gelman et al. [88] was adopted

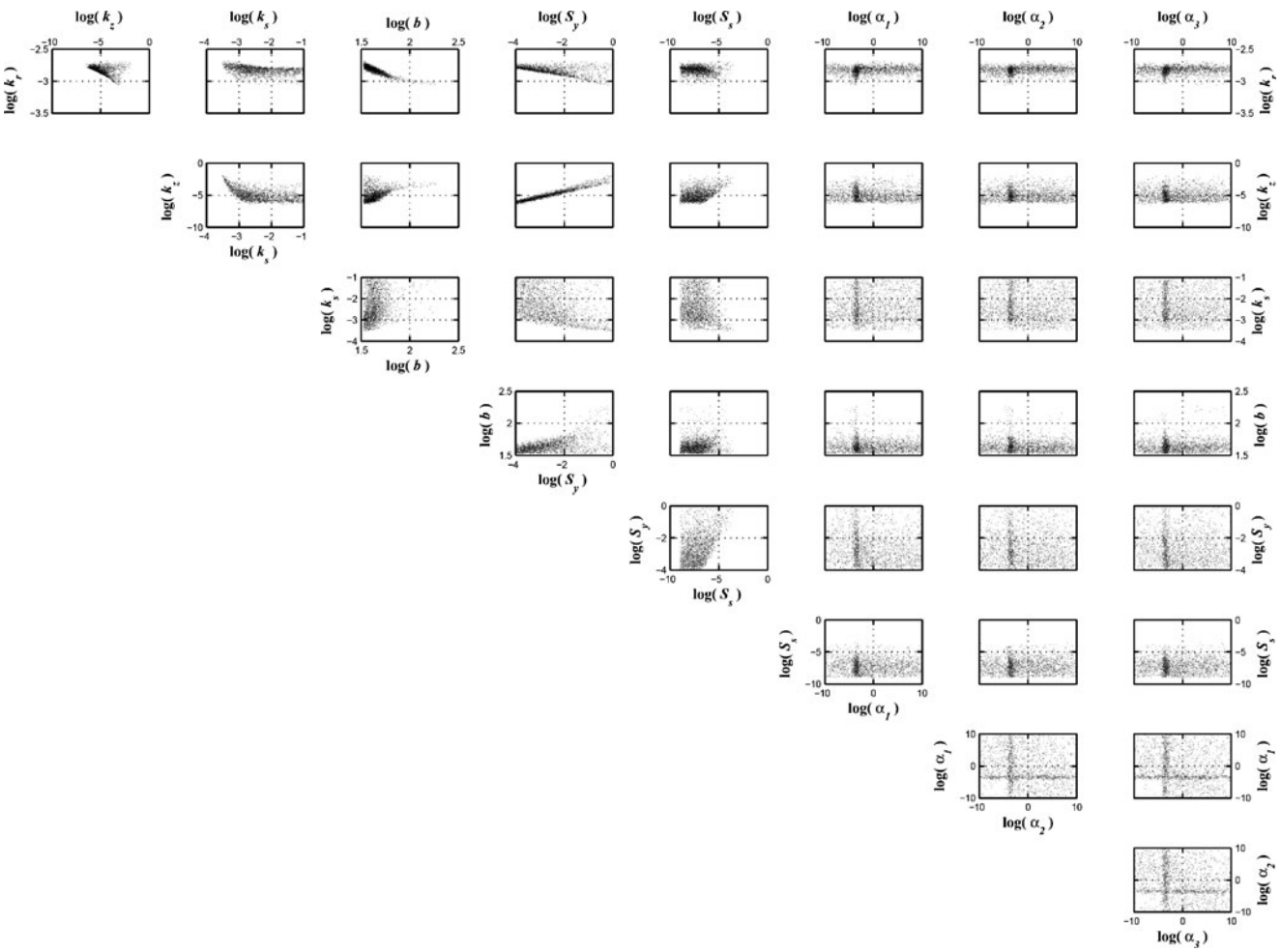

Fig. 3 Sample of the PDF of the pumping well

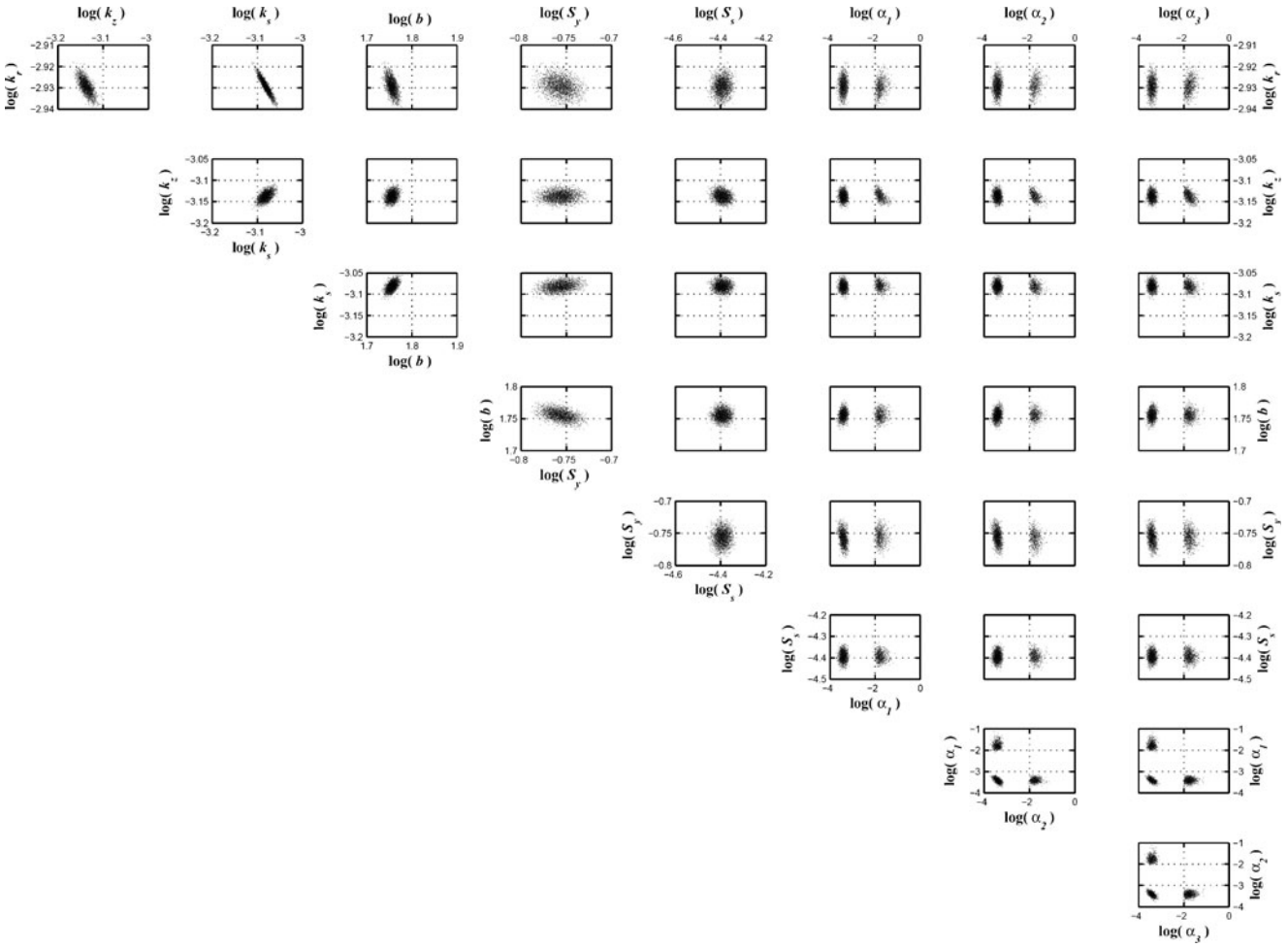

Fig. 4 Sample of the PDF of the joint problem

for obtaining a small initial sample of ten chains with 100 points each. The proposal function taken was a normal function with a diagonal correlation matrix (Fig. 2) and low values of the variance components. The chains cannot explore all the parameter space and consequently they get stuck. For that reason, the convergence, measured by the estimator R_{est} (potential scale reduction factor [88]), is not attained for all the parameters (Fig. 2). This process is restarted with $q(\mathbf{m}_i, \mathbf{m}_c) \sim N\left(\hat{\mathbf{m}}_s, \hat{\mathbf{C}}_s\right)$ where $\hat{\mathbf{m}}_s$ and $\hat{\mathbf{C}}_s$ are the mean and covariance matrix of the initial sample respectively and repeated until R_{est} and its upper bound, R_{ub}, are below 1.2 and near to 1 (Fig. 2). For the pumping well (F507080) and the joint problem, samples shown in Figs. 3 and 4 were obtained. In order to depict the sample, it is necessary to draw $n(n-1)/2$ two-dimensional scatter plots (in our example, 36 plots since the number of parameters to be identified is $n = 9$). These plots are a projection of the sampled points on the planes obtained after grouping the parameters in pairs, which is a graphical approximation of the marginal probability density. This type of representation shows the main trends of $p(\mathbf{m}|\mathbf{d})$. A comparison of the two figures shows that the uncertainty associated with the

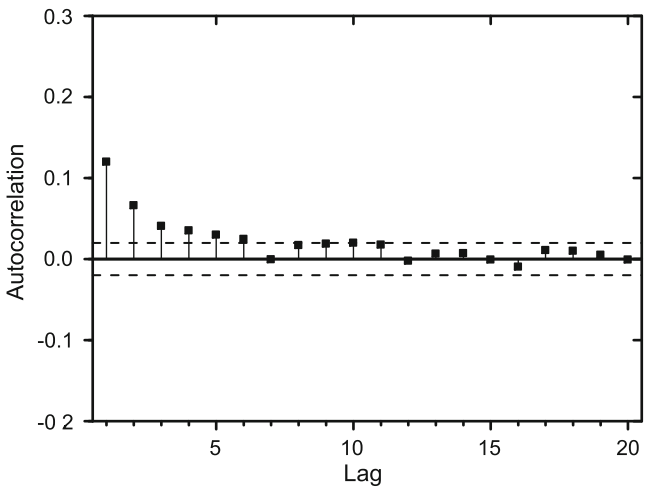

Fig. 5 Autocorrelogram for the pumped well (F507080). Autocorrelation falls below significant values after the sixth iteration

parameters of the pumping well is substantially greater than the uncertainty presented in the joint problem.

6.3 Marginal density estimation

On the basis of the MCMC sample shown in Figs. 3 and 4, and by applying the IWMDE or the KDE [102] (with which better results were had), the marginal PDF of each variable was computed. However, the high correlation between samples (a point usually is correlated with the next ones in the chain) caused by a poor mixing of the chains, can affect the inference based on them. For this reason the autocorrelogram is constructed (Fig. 5) and, in the case of the pumped well, there is significant correlation until lag 6. Consequently, only one every seven points of the total sample is selected for the computation of the marginal densities. This procedure for obtaining an independent sample by reducing the length of the chains is called "thinning". Furthermore, the first half of the chains is eliminated as burn-in [88]. With this reduced sample, the marginal PDFs are estimated. In the computation of the IWMDE the truncated multidimensional normal function was used as the weight function. The selected function minimizes the distances between this function and $p(\mathbf{m}|\mathbf{d})$. To make sure that all the marginal densities of $w(\mathbf{m})$ are normalized, the integrals of the truncated multidimensional normal function must be calculated [128]. In the case of the KDE, two alternatives were considered: a triangular kernel with an amplitude of one tenth of the range of the values and the Botev KDE [102]. In the case of the pumped well (Fig. 6), the IWMDE has yielded non-convergent results, probably caused by an inadequate selection of the weight function, which is one of the key points of this estimator. However, both KDE estimates were stable

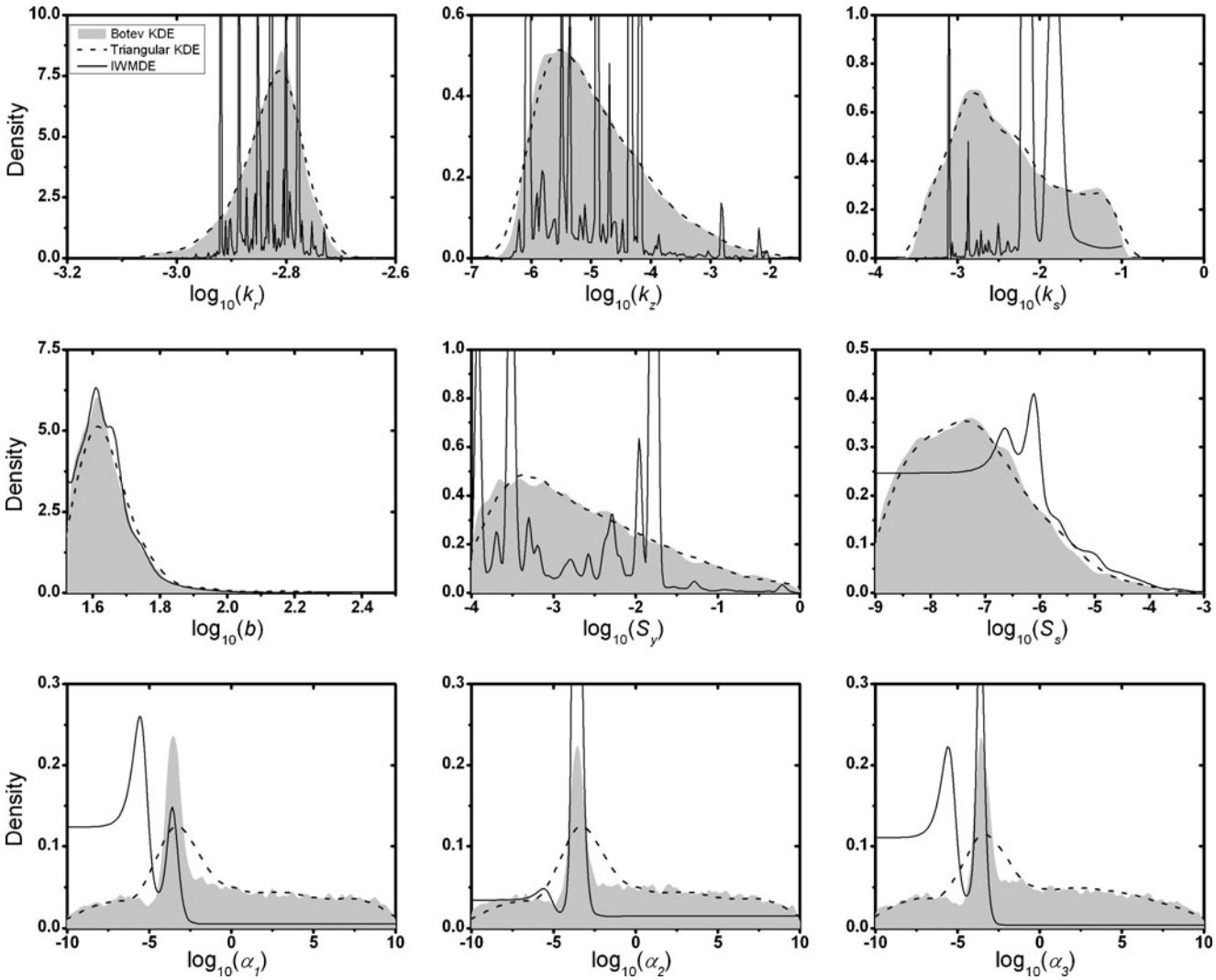

Fig. 6 Marginal density estimation for the pumped well (F507080)

and almost coincident, showing that these marginal density estimators are robust and relatively accurate. As can be seen in Figs. 6 and 7, the estimated marginal PDFs provide a reasonably coherent reproduction of the sample dispersion.

Moreover, the results of Fig. 7 show that, in some cases, they are far from being drawn from a normal density function (especially for the multimodal distributions of the α_i parameters). Consequently, in these cases, the classical confidence intervals may be of no use to the modeler. The highest probability intervals [58] are more appropriate for the marginal posterior probability densities that exhibit asymmetry. There are both complex [129] and simplified [130] algorithms to determine these intervals. However, it is usually enough to show graphical representations of the kind presented in Figs. 6 and 7 to have a rough idea of the highest probability intervals and their asymmetry. This enables to compare these marginal PDF with the confidence intervals obtained by Moench et al. [119]. As can be seen in Fig. 7, there are variables (particularly the permeabilities k_r, k_z and k_s) whose uncertainty is characterized fairly well on the basis of estimations using classical methods. Moreover, in some

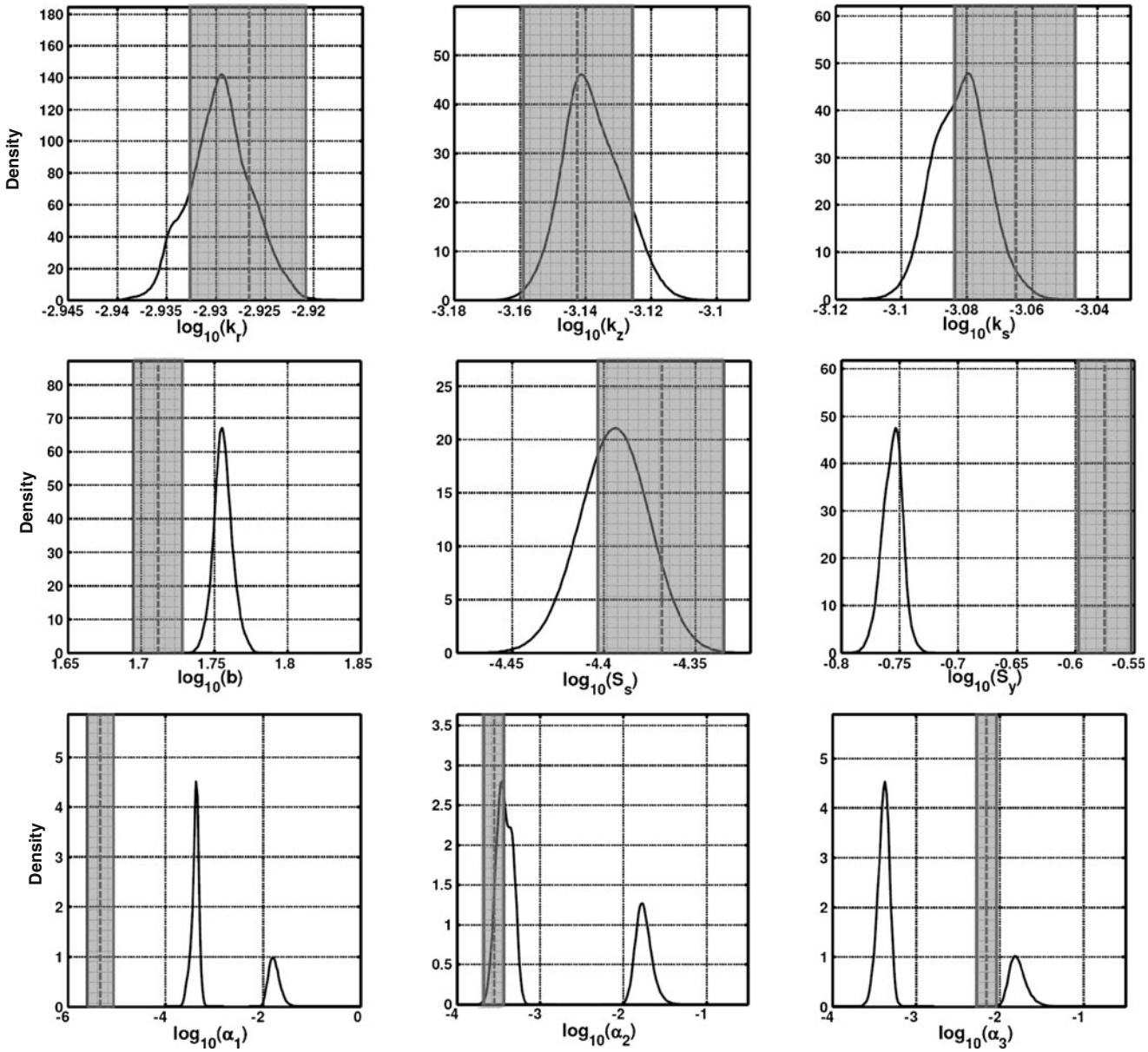

Fig. 7 Comparison of the marginal PDF of each parameter (*bold lines*), and the 95% confidence intervals (*grey shades*) obtained by Moench et al. [119]

Table 2 Marginal information of the observation wells and the pumped well

# well	Well id.	Marginal information (nats)									Total
		k_r(m/s)	k_z(m/s)	k_s(m/s)	b(m)	S_y(-)	S_s(m^{-1})	$\alpha_1(s^{-1})$	$\alpha_2(s^{-1})$	$\alpha_3(s^{-1})$	information
1	F347031	2.182	1.060	0.241	0.343	1.764	0.355	0.033	0.021	0.032	10.137
2	F376037	3.012	2.685	0.381	0.118	1.832	0.431	0.012	0.016	0.060	11.308
3	F377037	3.254	2.614	0.888	1.226	1.940	0.466	0.062	0.063	0.092	16.550
4	F381056	2.944	3.495	0.558	0.669	1.804	0.523	0.030	0.038	0.022	12.941
5	F383032	3.343	2.230	0.398	1.056	1.749	0.366	0.031	0.058	0.037	14.306
6	F383061	3.007	3.541	1.154	0.678	1.813	0.545	0.048	0.055	0.036	16.387
7	F383082	2.844	3.522	0.489	0.652	1.581	0.449	0.015	0.026	0.028	13.949
8	F383129	2.577	2.498	0.442	0.499	1.417	0.466	0.048	0.030	0.021	12.364
9	F384033	3.061	3.078	0.511	0.355	1.935	0.477	0.014	0.075	0.017	13.572
10	F385032	2.855	1.702	0.250	0.107	1.813	0.350	0.037	0.038	0.037	10.228
11	F434060	3.503	2.849	0.617	0.565	1.761	0.766	0.050	0.043	0.069	16.766
12	F450061	3.237	2.742	0.448	0.652	1.672	0.373	0.018	0.021	0.028	14.722
13	F476061	3.377	2.691	0.443	0.794	1.783	0.362	0.014	0.055	0.039	14.838
14	F478061	3.056	2.548	0.397	0.648	1.802	0.347	0.026	0.025	0.046	13.262
15	F504032	3.719	3.275	0.809	1.591	2.097	1.086	0.283	0.284	0.263	21.470
16	F504060	3.156	3.886	1.000	0.887	1.655	3.030	0.091	0.140	0.131	21.185
17	F504080	3.123	3.715	0.913	0.666	1.623	2.890	0.062	0.025	0.019	19.179
18	F505032	4.001	3.458	1.037	1.814	2.276	1.229	0.294	1.765	0.320	26.510
19	F505059	3.834	3.792	1.197	0.917	2.024	3.316	0.357	0.316	0.264	25.147
20	F505080	3.317	2.848	1.032	1.154	1.435	2.899	0.096	0.075	0.122	21.226
21	F507080	3.497	0.852	1.224	1.541	0.190	0.783	0.142	0.107	0.124	17.888
	All	6.265	5.205	5.263	3.829	4.443	4.579	2.535	2.675	2.596	44.807

of the parameters this determination is insufficient. This deficiency may be attributed to the multimodal character of the PDF and to the linearized approximation of the posterior covariance matrix around the estimated optimum.

6.4 Application of the information theory

If Shannon's information is evaluated for each piezometer and the joint problem, the results shown in Table 2 are obtained. This table presents the amount of information contained in the marginal density functions of each parameter as well as the joint posterior probability density function (PDF) of all of them. As evidenced by Fig. 8, the variables presenting greater uncertainty (Figs. 6 and 7) coincide with those containing less marginal information. Since the piezometers closest to the pumping well are more influenced by the vertical component of flow, it can be verified that there are groups of piezometers with a high information level that coincide with the groups proposed by Moench et al. [119], in keeping with their proximity to the pumped well. The closest piezometers (like F505032 and F505059) are able to contribute a very high quantity of global information as compared with the rest, since they are located in the area of the aquifer where the greatest variations in the pressure field are produced. On the other hand, the shallow, distant piezometers (like F347031, F376037 and F385032) present a low amount of global information, which implies a great deal of uncertainty in the estimation of the parameters by the classical optimization methods. Figure 8 shows the results of Table 2 in the form of a Box-and-Whisker diagram. The data pertaining to piezometer F505032

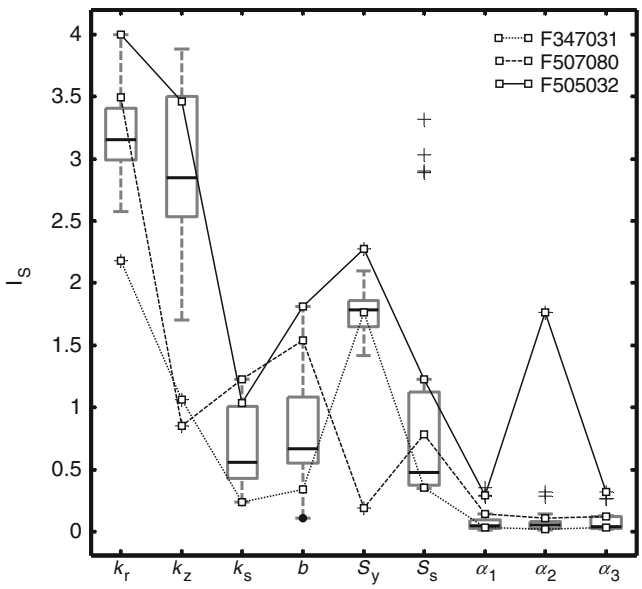

Fig. 8 Box plot of the marginal Shannon information for each parameter

(close to the pumping well), have been superimposed on piezometer F347031 (far from the pumping well) and on piezometer F507080 (test well). For F505032 all the marginal information values are above the median, while for F347031 it falls generally below the median. The marginal information values of the test well are similar to the values of the other two piezometers, except in the case of k_s, k_z and S_y. This is due to the high sensitivity of the drawdowns in the pumping well to the first parameter, and the low sensitivity to the latter two. As it could be expected, the marginal information of each variable in the complete problem is much higher than what was found for the observation wells on an individual basis. On the other hand, if we compare the data in Table 2 regarding global and marginal information of the complete problem, with the data obtained solely from the interpretation of the pumped well information (well #21, id. F507080, see Table 2), the difference is substantial. This is illustrated in Fig. 9, which shows the marginal densities, where it is possible to observe the differences inferred from the marginal information more clearly.

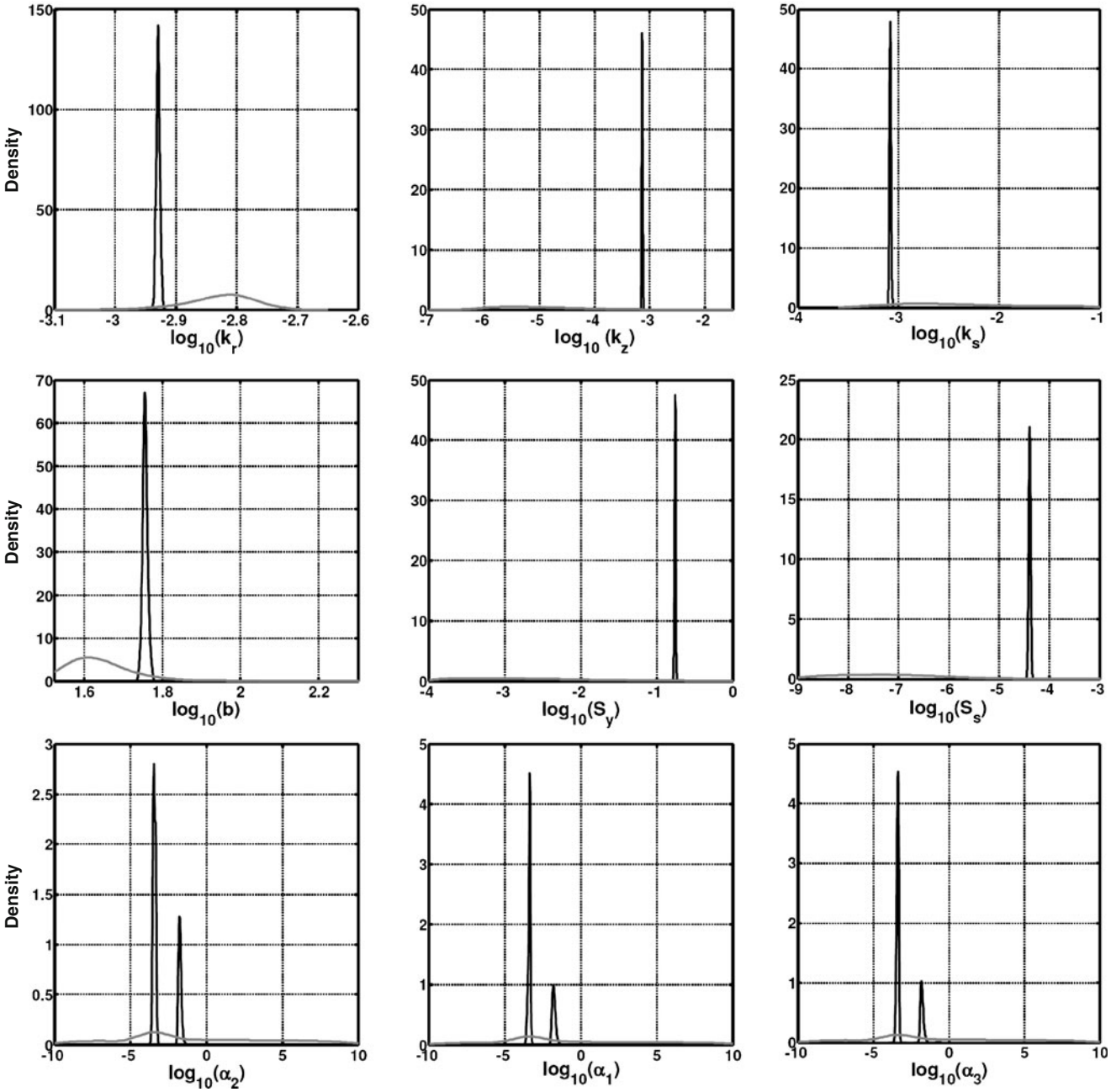

Fig. 9 Comparison of the marginal PDFs of the complete dataset (*bold line*) and the pumped well (*grey line*)

Table 3 Piezometer groups proposed

Group	Obs. wells
A	F505080 F504080 F383082 F383129
B	F505059 F504060 F383061
C	F505032 F504032 F377037 F383032
D	F384033 F381056 F347031 F385032 F376037
E	F434060 F450061 F476061 F478061
F	F505032 F505059 F505080
G	F504032 F504060 F504080
H	F383032 F383061 F383082 F383129
I	F504032 F504060 F505032 F505059 F505080
J	F504060 F505032 F505059 F507080

In terms of ill-posedness of the problem, the low marginal information of parameters such as S_y, or the set of α_i in the case of the pumping well (F507080), denotes a problem of non-identifiability (since a non-informative prior was selected). Although convergence of the MCMC sample was attained in this case (Fig. 2), it is only due to the fact that the space of the parameters is bounded as is shown in Table 1. Otherwise, an erratic behavior of the Metropolis–Hastings sampler would be expected [38]. Furthermore, as it can be observed in Fig. 3, S_y presents a strong correlation with k_z, which suggests that those variables cannot be identified in this single-well test (as suggested in [131]).

In a further analysis, Moench et al. [119] selected several piezometer groups in order to determine the effectiveness of different piezometer locations. This comparison was made on the basis of the values of the MLE. Given that ill-posedness is common to hydrogeologic problems, minimum changes in the dataset might lead to values that are very different from the estimated optimums. Consequently, the assessment of the suitability of the groups in terms of the distance between the MLEs may yield rather inconsistent results. Moench et al. [119] proposed eight possible combinations of the states of knowledge (Table 3, groups A to H). In order to evaluate the applicability of information theory for the design of experiments, the groups I and J are proposed based on the results shown in Table 2. In group I, the observation wells with global information greater than 20 nats were selected. Group J considers piezometers that present at least one marginal PDF with maximum information.

As expected, groups F and G contain a quantity of information that is close to the whole problem (see Table 4), since both the marginal and global information have very similar values. The proximity of the observation wells to the pumped well was found to determine the amount of global information that will be obtained after the pumping test interpretation. As far as groups I and J are concerned, the results are even closer to those of the complete data set, since they include most of the piezometers belonging to groups F and G. It is interesting to note that the presence of the pumped well proved to be fundamental to reducing the uncertainty of variable k_s in the case of J. The application of information theory for the selection of the optimal experimental setup has provided satisfactory results in this case. Although only three piezometers and the pumping well were used for the interpretation of the well, they yielded similar results to those obtained by the complete observational network (20 piezometers). This reduction of the number of control points highlights the importance in the selection of non-redundant monitoring networks for future model calibration.

The values corresponding to the marginal and total information confirm the intuitive assessment of the suitability of the groups in terms of the distance between MLEs made by Moench et al. [119]. For example, piezometer groups D, E and H contain substantially less information than piezometer groups F and G. The maximum likelihood approach, however, is not capable of classifying the groups either in terms of the

Table 4 Marginal information of the piezometer groups proposed

Cluster	Marginal information (nats)									Total
	k_r (m/s)	k_z (m/s)	k_s(m/s)	b(m)	S_y(-)	S_s(m^{-1})	$\alpha_1(s^{-1})$	$\alpha_2(s^{-1})$	$\alpha_3(s^{-1})$	information
A	5.410	4.330	1.303	3.104	3.280	4.120	0.509	0.494	0.683	28.039
B	5.621	4.817	1.309	2.921	2.923	4.133	0.304	0.239	0.170	28.789
C	5.932	4.347	1.197	2.882	3.725	3.451	0.951	0.997	0.862	30.848
D	3.927	3.358	0.575	1.628	2.319	0.520	0.055	0.047	0.049	17.083
E	5.068	3.614	2.172	2.264	2.406	0.383	0.088	0.081	0.075	21.285
F	4.673	4.252	2.276	2.547	2.817	3.212	1.802	1.774	1.370	36.277
G	4.510	4.492	1.089	2.272	3.164	3.892	2.443	2.462	2.412	32.323
H	3.977	3.736	0.900	2.528	2.305	0.899	0.135	0.118	0.143	21.402
I	5.798	5.085	1.590	3.373	3.815	3.741	2.583	2.028	2.694	37.295
J	5.665	5.064	4.899	3.181	3.582	4.382	2.695	2.688	2.495	40.763
All	6.265	5.205	5.263	3.829	4.443	4.579	2.535	2.675	2.596	44.807

Table 5 Posterior diagnostics of the assumption of the normality distribution of the residuals

Statistic	Value of statistic	5% significance critical value	Reject normality?
Kolmogorov–Smirnov	4.85×10^{-2}	6.13×10^{-2}	No

information provided or by the uncertainty in the evaluation of its parameters on an individual basis. Neither is this approach able to facilitate the task of grouping the piezometers according to objective criteria. Piezometer groups I and J, established on the basis of information theory, provide much more information than any of the other groups proposed.

6.5 Posterior diagnostics

The validity of the assumptions made in the construction of the posterior PDF should be tested. Therefore, in order to be consistent with the diagnostics proposed above, several features of the posterior sample are examined. Firstly the comparison of the parameter estimates obtained by Moench [119] and the posterior marginals shows similar features in most of the parameters (Fig. 7). However, some other parameters such as S_y and α_i, present discrepancies but these can be attributed to the multimodality of the posterior PDF.

The normality of the residuals was examined using the Kolmogorov–Smirnov [132] test. In this case, the null hypothesis that the residuals follow a normal distribution cannot be rejected (Table 5). The normal probability plot [110] of the residuals (Fig. 10) also suggests a Gaussian distribution of the errors.

Finally the posterior PDF of the joint problem does not exhibit ill-posedness, as can be observed in Table 2, where all marginal information values are higher than 2.5 nats. Furthermore, the convergence of the MCMC sampler (Fig. 2) suggests the same conclusion. Thus, ill-posedness can be disregarded in the present problem.

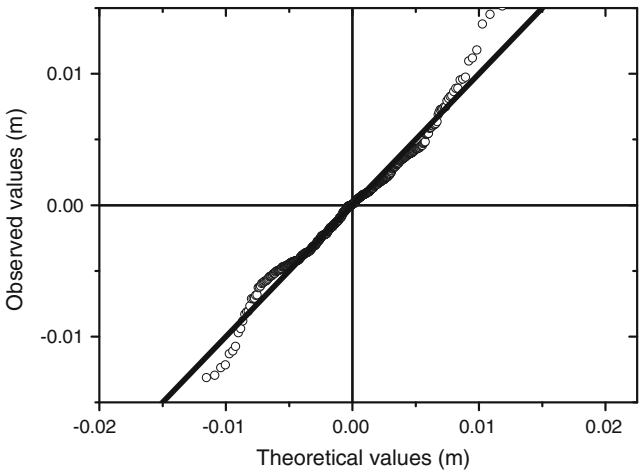

Fig. 10 Normal probability plot of the residualsl

7 Concluding remarks

This paper has recommended a methodology for the analysis of inverse problems, based on three basic concepts: the Bayesian approach to inverse problems, the Markov Chain Monte Carlo algorithms and Shannon's information. Although they are all well known and have been applied individually, they have been used in combination only occasionally in the field of geosciences.

Firstly, a collection of algorithms based on MCMC methods for inverse problem analysis based on Bayesian statistics has been reviewed in this paper. The Bayesian approach to inverse problem analysis has been applied to a variety of problems in hydrological sciences. However, prior to the emergence of the MCMC methods, this approach was not common in complex problems without closed form descriptors of the posterior density functions (such as their mode, mean or variance). Despite of its versatility, the high computational costs of this kind of analysis may not be very competitive as compared to those based on the optimization of the posterior density of the parameters. However, the degree of knowledge acquired about the problem in question makes it especially interesting in research or in problems where the uncertainty quantification is the central issue. Nevertheless, the correct construction of the likelihood function, the identification of the sources of uncertainty and the posterior diagnostics of the assumptions made are essential if a meaningful analysis is to be carried out.

The Shannon's information measure is the second basic tool within this approach for inverse problem analysis. The importance of this measure lies in its global character, based on the integration of the posterior density of the parameters in the entire domain. This enables the modeler to evaluate how much incremental information content comes from each data value, and to compare different experimental set-ups. Furthermore, an evaluation of the ill-posedness (or non-identifiability) of the problem can be done by the examination of the marginal information content of each parameter. Although this diagnostic tool is not conclusive, it can be used together with other methods such as the non-convergence of the MCMC samplers or the condition number of the Hessian matrix at the maximum likelihood estimator.

Despite their high computational expense, the MCMC methods are competitive as compared to other integration or sampling methods in highly multidimensional spaces. Although there are other alternatives to reduce the total calculation time such as parallelization (MCMC algorithms are easily parallelizable, especially if evolutionary algorithms are used as proposal functions of the Metropolis–Hastings sampler) or the improvement of the algorithm that resolves the direct problem considered. In the present paper, several computationally inexpensive estimators have been presented in order to reduce the computational effort to the sampling procedure.

Finally the case study presented shows how the MCMC methods can be applied in a common inverse problem: a pumping test. The uncertainty of the parameters has been characterized, as well as the total and marginal information content of the whole problem. The results obtained are more accurate that the ones found in the literature [119] computed using the maximum likelihood approach, mainly due to the multimodal posterior PDF. Furthermore, an experimental set-up that has reduced the total number of piezometers to be used from 21 to 4 has been proposed. In the end, an evaluation of the adequacy of the hypothesis assumed during the construction of the posterior PDF was carried out. The results confirm the validity of the assumptions made and, consequently, the validity of the uncertainty analysis carried in this case study.

On the whole, this approach offers a consistent way of analyzing inverse problems, with a higher computational cost, but with a higher degree of knowledge on the uncertainty of the parameters and the total information content of the observed data.

Acknowledgements This research was financed in part by a grant awarded to the first author by the Education and Research Department of the Castilla-La Mancha Regional Government and the European Social Fund within the framework of the Integrated Operative Programme for Castilla-La Mancha 2000–2006, approved by Commission Decision C(2001) 525/1. This research was also financed in part by the Spanish Ministry of Education through the FPU Grant AP2009–2134 awarded to the second author and by the Research Grants of the Spanish Ministry of Science and Innovation PSS-120000–2007–27 and PSS-120000–2008–31 awarded to the third author.

References

1. Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J.: Inverse problem in hydrogeology. Hydrogeol. J. **13**(1), 206–222 (2005)
2. Burnham, K.P., Anderson, D.R.: Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer, New York (2002)
3. Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philos. trans. R. Soc. Lond., A **222**, 309–368 (1922)
4. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Automat. Contr. **AC-19**(6), 716–723 (1974)
5. de Leeuw, J.: Information theory and an extension of the maximum likelihood principle by Hirotogu Akaike. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics, pp. 599–609. Springer, Berlin (1990)
6. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
7. Rissanen, J.: Stochastic complexity. J. R. Stat. Soc., B **49**(3), 223–239 (1987)
8. Kashyap, R.L.: Optimal choice of Ar and Ma parts in augoregressive moving average models. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-4**(2), 99–104 (1982)
9. Young, P., Wallis, S.: Recursive estimation: a unified approach to the identification estimation, and forecasting of hydrological systems. Appl. Math. Comput. **17**(4), 299–334 (1985)
10. Young, P.C., Beven, K.J.: Data-based mechanistic modelling and the rainfall-flow non-linearity. Environmetrics **5**(3), 335–363 (1994)
11. Neuman, S.P., Carrera, J.: Maximum-likelihood adjoint-state finite-element estimation of groundwater parameters under steady- and nonsteady-state conditions. Appl. Math. Comput. **17**(4), 405–432 (1985)
12. Noakes, D.J., McLeod, A.I., Hipel, K.W.: Forecasting monthly riverflow time series. Int. J. Forecast. **1**(2), 179–190 (1985)
13. Ledesma, A., Gens, A., Alonso, E.E.: Parameter and variance estimation in geotechnical backanalysis using prior information. Int. J. Numer. Anal. Methods Geomech. **20**(2), 119–141 (1996)
14. Di Baldassarre, G., Laio, F., Montanari, A.: Design flood estimation using model selection criteria. Phys. Chem. Earth, A/B/C **34**(10–12), 606–611 (2009)
15. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1985)
16. Diciccio, T.J., Kass, R.E., Raftery, A., Wasserman, L.: Computing Bayes factors by combining simulation and asymptotic approximations. J. Am. Stat. Assoc. **92**(439), 903–915 (1997)
17. Kass, R.E., Raftery, A.E.: Bayes factors. J. Am. Stat. Assoc. **90**(430), 773–895 (1995)
18. Perreault, L., Bernier, J., Bobée, B., Parent, E.: Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. J. Hydrol. **235**(3–4), 242–263 (2000)
19. Poeter, E., Anderson, D.: Multimodel ranking and inference in ground water modeling. Ground Water **43**(4), 597–605 (2005)
20. Marshall, L., Nott, D., Sharma, A.: Hydrological model selection: a Bayesian alternative. Water Resour. Res. **41**(10), W10422 (2005)
21. Frost, A.J., Thyer, M.A., Srikanthan, R., Kuczera, G.: A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data. J. Hydrol. **340**(3–4), 129–148 (2007)
22. Neuman, S.P.: Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch. Environ. Res. Risk Assess. **17**(5), 291–305 (2003)
23. Ajami, N.K., Duan, Q., Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty

in hydrologic prediction. Water Resour. Res. **43**(1), W01403 (2007)
24. Carrera, J., Neuman, S.P.: Estimation of aquifer parameters under transient and steady state conditions: I. Maximum likelihood method incorporating prior information. Water Resour. Res. **22**(2), 199–210 (1986)
25. Carrera, J., Neuman, S.P.: Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. Water Resour. Res. **22**(2), 211–227 (1986)
26. McLaughlin, D., Townley, L.R.: A reassessment of the groundwater inverse problem. Water Resour. Res. **32**(5), 1131–1161 (1996)
27. Young, P.: Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis. Comput. Phys. Commun. **117**(1–2), 113–129 (1999)
28. Young, P.C., McKenna, P., Bruun, J.: Identification of nonlinear stochastic systems by state dependent parameter estimation. Int. J. Control **74**(18), 1837–1857 (2001)
29. Young, P.C.: The data-based mechanistic approach to the modelling, forecasting and control of environmental systems. Annu. Rev. Control **30**(2), 169–182 (2006)
30. Kuczera, G., Kavetski, D., Franks, S., Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. J. Hydrol. **331**(1–2), 161–177 (2006)
31. Romanowicz, R.J., Young, P.C., Beven, K.J.: Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom. Water Resour. Res. **42**(6), W06407 (2006)
32. Reichert, P., Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour. Res. **45**(10), W10402 (2009)
33. Aldrich, J.: R. A. Fisher and the making of maximum likelihood 1912–1922. Stat. Sci. **12**(3), 162–176 (1997)
34. Edwards, A.W.F.: Likelihood, Expanded edn. Johns Hopkins University Press, Baltimore (1992)
35. Kay, S.M.: Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory. Prentice Hall, Upper Saddle River (1993)
36. Yeh, T.-C.J., Simunek, J.: Stochastic fusion of information for characterizing and monitoring the vadose zone. Vadose Zone J. **1**(2), 207–221 (2002)
37. Yeh, T.C.J., Lee, C.H., Hsu, K.C., Tan, Y.C.: Fusion of active and passive hydrologic and geophysical tomographic surveys: the future of subsurface characterization. Data integration in subsurface hydrology. Geophys. Monogr. **171**, 109–120 (2007)
38. Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W.: Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. **46**(5), W05521 (2010)
39. Gelfand, A.E., Sahu, S.K.: Identifiability, improper priors and Gibbs sampling for generalized linear models. J. Am. Stat. Assoc. **94**(445), 247–253 (1999)
40. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Wiley, New York (1977)
41. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems. Academic Press, San Diego (2005)
42. Howie, D.: Interpreting Probability: Controversies and Developments in the Early Twentieth Century. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge University Press, Cambridge (2002)
43. Jaynes, E.T., Bretthorst, G.L.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge (2003)
44. Jeffreys, H.: Theory of Probability, 3rd edn. Oxford Classic Texts in the Physical Sciences. Oxford University Press, Oxford (1998)
45. Jaynes, E.T.: How does the brain do plausible reasoning? In: Erickson, G.J., Smith, C.R. (eds.) Maximum-Entropy and Bayesian Methods in Science and Engineering, vol. 1, pp. 1–24. Kluwer, Dordrecht (1988)
46. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia, USA (2005)
47. Tarantola, A., Valette, B.: Generalized nonlinear inverse problems solved using the least squares criterion. Rev. Geophys. Space Phys. **20**(2), 219–232 (1982)
48. Hanna, S., Yeh, T.C.J.: Estimation of co-conditional moments of transmissivity, hydraulic head, and velocity fields. Adv. Water Resour. **22**(1), 87–95 (1998)
49. Lehman, E.L., Casella, G.: Theory of Point Estimation, 2nd edn. Springer, New York (1998)
50. Woodbury, A., Sudicky, E., Ulrych, T.J., Ludwig, R.: Three-dimensional plume source reconstruction using minimum relative entropy inversion. J. Contam. Hydrol. **32**(1–2), 131–158 (1998)
51. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)
52. Vogel, C.R.: Computational methods for inverse problems. Frontiers in Applied Mathematics. SIAM, University City Science, Center, PATraduce (2002)
53. Zhdanov, M.: Geophysical inverse theory and regularization problems, vol. 36. Methods in Geochemistry and Geophysics. Elsevier, Amsterdam (2002)
54. Mosegaard, K., Tarantola, A.: Probabilistic approach to inverse problems. In: Lee, W.H.K., Kanamori, H., Jennings, P., Kisslinger, C. (eds.) International Handbook of Earthquake and Engineering Seismology, vol. A, pp. 237–265. Academic Press, London (2002)
55. Skahill, B.E., Doherty, J.: Efficient accommodation of local minima in watershed model calibration. J. Hydrol. **329**(1–2), 122–139 (2006)
56. Kavetski, D., Clark, M.P.: Ancient numerical Daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resour. Res. **46**(10), W10511 (2010)
57. Beven, K., Binley, A.: The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. **6**(3), 279–298 (1992)
58. Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis, Reprint edn. Wiley, New York (1992)
59. Ulrych, T.J., Sacchi, M.D., Woodbury, A.: A Bayes tour of inversion: a tutorial. Geophysics **66**(1), 55–69 (2001)
60. Hadamard, J.: Sur les problèmes aux dérivées partielles et leur signification physique, pp. 49–52. Princeton University Bulletin (1902)
61. Bennet, J.H. (ed.) R. A. Fisher. Statistical Inference and Analysis. Selected correspondence of R.A. Fisher. Oxford University Press, Oxford (1990)
62. Kavetski, D., Kuczera, G., Franks, S.W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory - art. no. W03407. Water Resour. Res. **42**(3), 3407–3407 (2006)
63. Kavetski, D., Kuczera, G., Franks, S.W.: Bayesian analysis of input uncertainty in hydrological modeling: 2.

Application - art. no. W03408. Water Resour. Res. **42**(3), 3408–3408 (2006)
64. Bretthorst, G.L.: An Introduction to Parameter-Estimation Using Bayesian Probability-Theory. In: Fougère, P.F. (ed.) Maximum Entropy and Bayesian Methods ///, vol. 39. Fundamental Theories of Physics, pp. 53–79. Kluwer, Dordrecht (1990)
65. Amar, J.G.: The Monte Carlo method in science and engineering. Comput. Sci. Eng. **8**(2), 9–19 (2006)
66. Beichl, I., Sullivan, F.: Monte Carlo methods. Comput. Sci. Eng. **8**(2), 7–8 (2006)
67. Sorensen, D.: Likelihood. Bayesian and MCMC Methods in Genetics. Springer, New York (2002)
68. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**(6), 1087–1092 (1953)
69. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97–109 (1970)
70. Kuczera, G., Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J. Hydrol. **211**(1–4), 69–85 (1998)
71. Vrugt, J.A., Bouten, W., Gupta, H.V., Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: the information content of experimental data. Water Resour. Res. **38**(12), 481–4813 (2002)
72. Fu, J., Jaime Gómez-Hernández, J.: Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov Chain Monte Carlo method. J. Hydrol. **364**(3–4), 328–341 (2009)
73. Chen, J., Hubbard, S., Rubin, Y., Murray, C., Roden, E., Majer, E.: Geochemical characterization using geophysical data and Markov Chain Monte Carlo methods: a case study at the South Oyster bacterial transport site in Virginia. Water Resour. Res. **40**(12), 1–14 (2004)
74. Chen, J., Hubbard, S., Peterson, J., Williams, K., Fienen, M., Jardine, P., Watson, D.: Development of a joint hydrogeophysical inversion approach and application to a contaminated fractured aquifer. Water Resour. Res. **42**(6), W06425 (2006)
75. Fu, J., Gómez-Hernández, J.J.: A blocking Markov Chain Monte Carlo method for inverse stochastic hydrogeological modeling. Math. Geosci. **41**(2), 105–128 (2009)
76. Oliver, D.S., Cunha, L.B., Reynolds, A.C.: Markov Chain Monte Carlo methods for conditioning a permeability field to pressure data. Math. Geol. **29**(1), 61–91 (1997)
77. Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X., Mallick, B.: An efficient two-stage Markov Chain Monte Carlo method for dynamic data integration. Water Resour. Res. **41**(12), 1–6 (2005)
78. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**(6), 721–741 (1984)
79. Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M.: Illustration of Bayesian inference in normal data models using gibbs sampling. J. Am. Stat. Assoc. **85**(412), 972–985 (1990)
80. Kuczera, G., Kavetski, D., Renard, B., Thyer, M.: A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. Water Resour. Res. **46**(7), W07602 (2010)
81. Baragona, R., Battaglia, F., Poli, I.: Evolutionary Statistical Procedures. Statistics and Computing. Springer, Heidelberg (2011)
82. Drugan, M.M., Thierens, D.: Evolutionary Markov Chain Monte Carlo. Lect. Notes Comput. Sci. **2936**, 63–76 (2004)
83. Ter Braak, C.J.F.: A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. Stat. Comput. **16**(3), 239–249 (2006)
84. Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D.: Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. of Nonlinear Sci. & Numer. Simul. **10**(3), 273–290 (2009)
85. Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Environ. Res. Risk Assess. **23**(7), 1011–1026 (2009)
86. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Comput. Geosci. **15**(1), 185–221 (2011)
87. Brooks, S.P., Gelman, A.: General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. **7**(4), 434–455 (1998)
88. Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4), 457–472 (1992)
89. Sorensen, D.A., Andersen, S., Gianola, D., Korsgaard, I.: Bayesian inference in threshold models using Gibbs sampling. Genet. Sel. Evol. **27**, 229–249 (1995)
90. Cowles, M.K., Carlin, B.P.: Markov Chain Monte Carlo convergence diagnostics: a comparative review. J. Am. Stat. Assoc. **91**(434), 883–904 (1996)
91. Chen, M.H., Shao, Q.M., Ibrahim, J.G.: Monte Carlo Methods in Bayesian Computation. Springer Series in Statistics, Springer, New York (2000)
92. Gamerman, D., Freitas Lopes, H.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC, Boca Raton (2006)
93. Efendiev, Y., Hou, T., Luo, W.: Preconditioning Markov Chain Monte Carlo simulations using coarse-scale models. SIAM J. Sci. Comput. **28**(2), 776–803 (2006)
94. Ginting, V., Pereira, F., Presho, M., Wo, S.: Application of the two-stage Markov Chain Monte Carlo method for characterization of fractured reservoirs using a surrogate flow model. Comput. Geosci. 1–17 (2011). doi:10.1007/s10596-011-9236-4
95. Emerick, A.A., Reynolds, A.C.: Combining the Ensemble Kalman Filter with Markov Chain Monte Carlo for Improved History Matching and Uncertainty Characterization. Paper presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 21–23 February (2011)
96. Geyer, C.J.: Practical Markov Chain Monte Carlo. Stat. Sci. **7**(4), 473–511 (1992)
97. Gelfand, A.E., Smith, A.F.M., Lee, T.M.: Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. J. Am. Stat. Assoc. **87**, 523–532 (1992)
98. Chen, M.-H.: Importance-weighted marginal Bayesian posterior density estimation. J. Am. Stat. Assoc. **89**(427), 818–824 (1994)
99. Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley Series in Probability and Statistics. Wiley, New York (1992)
100. Scott, D.W., Sain, S.R.: Multi-dimensional density estimation. In: Rao, C.R., Wegman, E.J., Solka, J. (eds.) Handbook of Statistics: 24 Data Mining and Computational Statistics, Data Mining and Computational Statistics, vol. 24, pp. 229–262. Elsevier, Amsterdam (2004)

101. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
102. Botev, Z.I., Grotowski, J.F., Kroese, D.P.: Kernel density estimation via diffusion. Ann. Stat. **38**(5), 2916–2957 (2010)
103. Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T.: Bayesian model averaging: a tutorial. Stat. Sci. **14**(4), 382–417 (1999)
104. Stewart, L.: Hierarchical bayesian analysis using Monte Carlo integration: computing posterior distributions when there are many possible models. Statistician **36**(2/3), 211–219 (1987)
105. Hammersley, J.M., Handscomb, D.C.: Monte Carlo Methods. Methuen's Monographs on Applied Probability and Statistics. Methuen & Co. LTD., London, UK (1964)
106. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in Fortran 77. The Art of Scientific Computing, vol. 1, 2nd edn. Cambridge University Press, Cambridge (1992)
107. Chen, M.H.: Computing marginal likelihoods from a single MCMC output. Stat. Neerl. **59**(1), 16–29 (2005)
108. Kavetski, D.: Introduction to Bayesian inference using standard and hierarchical models. Paper presented at the Systemanalyse und Modellierung, Bayesian Uncertainty Analysis, Viena, Austria (2008)
109. Thyer, M., Leonard, M., Kavetski, D., Need, S., Renard, B.: The open source RFortran library for accessing R from Fortran, with applications in environmental modelling. Environ. Model. Softw. **26**(2), 219–234 (2011)
110. Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A.: Graphical Methods for Data Analysis. Statistics/Probability Series. Wadsworth & Brooks/Cole, Pacific Grove, California, USA (1983)
111. Finsterle, S., Faybishenko, B.: Inverse modeling of a radial multistep outflow experiment for determining unsaturated hydraulic properties. Adv. Water Resour. **22**(5), 431–444 (1999)
112. De Pauw, D.J.W., Sin, G., Insel, G., Van Hulle, S.W.H., Vandenberghe, V., Vanrolleghem, P.A.: Discussion of "Assessing parameter identifiability of activated sludge model number 1" by Pedro Afonso and Maria da Conceição Cunha. J. Environ. Eng. **130**(1), 110–112 (2004)
113. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
114. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
115. Thyer, M., Frost, A.J., Kuczera, G.: Parameter estimation and model identification for stochastic models of annual hydrological data: is the observed record long enough? J. Hydrol. **330**(1–2), 313–328 (2006)
116. Gelfand, A.E., Dey, D.K.: Bayesian model choice: asymptotics and exact calculations. J. R. Stat. Soc., B **56**(3), 501–514 (1994)
117. Abellan, A., Noetinger, B.: Optimizing subsurface field data acquisition using information theory. Math. Geosci. **42**(6), 603–630 (2010)
118. Wu, C.M., Yeh, T.C.J., Zhu, J.F., Lee, T.H., Hsu, N.S., Chen, C.H., Sancho, A.F.: Traditional analysis of aquifer tests: comparing apples to oranges? Water Resour. Res. **41**(9), W09402 (2005)
119. Moench, A.F., Garabedian, S.P., LeBlanc, D.R.: Estimation of hydraulic parameters from an unconfined aquifer test conducted in a glacial outwash deposit, Cape Cod, Massachusetts. In: US Geological Survey Professional Paper, vol. 1629, pp. 1–51. US Geological Survey, Reston, Virginia (2001)
120. Moench, A.F.: Estimation of hectare-scale soil-moisture characteristics from aquifer-test data. J. Hydrol. **281**(1–2), 82–95 (2003)
121. Moench, A.F.: Importance of the vadose zone in analyses of unconfined aquifer tests. Ground Water **42**(2), 223–233 (2004)
122. Moench, A.F.: Combining the Neuman and Boulton models for flow to a well in an unconfined aquifer. Ground Water **33**(3), 378–384 (1995)
123. Moench, A.F.: Flow to a well of finite diameter in a homogeneous, anisotropic water table aquifer. Water Resour. Res. **33**(6), 1397–1407 (1997)
124. Leblanc, D.R.: Sewage plume in a sand and gravel aquifer, Cape Cod, Massachusetts. In: U.S. Geological Survey, Open-File Report 84-475, p. 28. U.S. Geological Survey, Washington (1984)
125. Leblanc, D.R., Guswa, J.H., Frimpter, M.H., Londquist, C.J.: Ground-water resources of Cape Cod, Massachusetts. In: U.S. Geological Survey, Hydrologic Atlas 692, p. 4. U.S. Geological Survey, Reston, Virginia (1986)
126. Masterson, J.P., Stone, B.D., Walters, D.A., Savoie, J.: Hydrogeologic framework of western Cape Cod, Massachusetts. In: U.S. Geological Survey, Hydrologic Atlas 741, p. 1. U.S. Geological Survey, Reston, Virginia (1997)
127. Doherty, J., Brebber, L., Whyte, P.: PEST: Model Independent Parameter Estimation, p. 249. Watermark Numerical Computing (1994)
128. Genz, A.: Numerical computation of multivariate normal probabilities. J. Comput. Graph. Stat. **1**(1), 141–149 (1992)
129. Tanner, M.A.: Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. Springer Series in Statistics. Springer, New York (1996)
130. Chen, M.H., Shao, Q.M.: Monte carlo estimation of bayesian credible and HPD interval. J. Comput. Graph. Stat. **8**(1), 69–92 (1999)
131. Boonstra, J., Kselik, R.A.L.: SATEM 2002: Software for Aquifer Test Evaluation. Publication 57. p. 148. International Institute for Land Reclamation and Improvement. Wageningen, The Netherlands (2001)
132. Massey, F.J.: The Kolmogorov–Smirnov test for goodness of fit. J. Am. Stat. Assoc. **46**(253), 68–78 (1951)