Published in Water Resources Research 46, issue W11530, 1-17, 2010
which should be used for any reference to this work


# Bayesian inverse problem and optimization with iterative spatial resampling

Grégoire Mariethoz,[1,2,3] Philippe Renard,[1] and Jef Caers[2]

[1] Measurements are often unable to uniquely characterize the subsurface at a desired modeling resolution. In particular, inverse problems involving the characterization of hydraulic properties are typically ill-posed since they generally present more unknowns than data. In a Bayesian context, solutions to such problems consist of a posterior ensemble of models that fit the data (up to a certain precision specified by a likelihood function) and that are a subset of a prior distribution. Two possible approaches for this problem are Markov chain Monte Carlo (McMC) techniques and optimization (calibration) methods. Both frameworks rely on a perturbation mechanism to steer the search for solutions. When the model parameters are spatially dependent variable fields obtained using geostatistical realizations, such as hydraulic conductivity or porosity, it is not trivial to incur perturbations that respect the prior spatial model. To overcome this problem, we propose a general transition kernel (iterative spatial resampling, ISR) that preserves any spatial model produced by conditional simulation. We also present a stochastic stopping criterion for the optimizations inspired from importance sampling. In the studied cases, this yields posterior distributions reasonably close to the ones obtained by a rejection sampler, but with a greatly reduced number of forward model runs. The technique is general in the sense that it can be used with any conditional geostatistical simulation method, whether it generates continuous or discrete variables. Therefore it allows sampling of different priors and conditioning to a variety of data types. Several examples are provided based on either multi-Gaussian or multiple-point statistics.

## 1. Introduction

[2] Integrating state variables in hydrogeological site characterization by solving an inverse problem continues to be an important topic of investigation [*Carrera et al.*, 2005; *Hendricks-Franssen et al.*, 2009; *Liu et al.*, 2010; *Zimmerman et al.*, 1998]. Indeed, inverse problems are a crucial aspect of groundwater modeling since they are used to validate or invalidate certain geological scenarios [*Ronayne et al.*, 2008] as well as to reduce model uncertainty for engineering prediction and decision making problems [*Alcolea et al.*, 2009]. As such, models need to not just match the data, they also need to be predictive, a property that is difficult to objectively verify [*Ballin et al.*, 1993; *Subbey et al.*, 2004]. Conditioning models to points data (localized measurements of the variable of interest) is addressed very efficiently by most geostatistical simulation algorithms [*Deutsch and Journel*, 1992; *Remy et al.*, 2009]. In this paper, we refer to conditioning models to indirect state variable data (such as heads).

[3] Problems involving flow in underground media typically present more unknowns than data. For example, modeling hydraulic conductivity or porosity on an entire domain, based only on local head measurements or tracer tests, is typically an ill-posed inverse problem [*Carrera et al.*, 2005; *De Marsily et al.*, 2005; *Yeh*, 1986]. Ill-posedness means that multiple solutions are possible, and characterizing the uncertainty spanned by these multiple solutions is often critical in real field engineering use of these models. Other consequences of ill-posedness can be that a solution does not exist or is instable with regard to small variations in the input data [*Carrera and Neuman*, 1986]. In a Bayesian framework, these issues are dealt with by obtaining a posterior distribution given a certain prior distribution and a likelihood function. In this respect, only Markov chain Monte Carlo (McMC) methods have been shown to sample with reasonable accuracy from this posterior [*Mosegaard and Tarantola*, 1995; *Omre and Tjelmeland*, 1996], i.e., to generate model realizations that (1) match the points data and the indirect state data, (2) reproduce for each inverse solution some prior statistics (e.g., a spatial covariance) and (3) sample correctly from the posterior as imposed by Bayes' rule. Most gradient-based/optimization techniques [*De Marsily et al.*, 1984; *Gomez-Hernandez et al.*, 1997; *Hernandez et al.*, 2006; *RamaRao et al.*, 1995; *Vesselinov et al.*, 2001] do not completely fulfill these three requirements.

[1] Centre for Hydrogeology, University of Neuchâtel, Neuchâtel, Switzerland.
[2] ERE Department, Stanford University, Stanford, California, USA.
[3] National Centre for Groundwater Research and Training, The University of New South Wales, Sydney, New South Wales, Australia.

[4] However, in many real-case problems, geostatistical simulations and evaluations of the forward problem are so CPU demanding that traditional McMC methods are not applicable. Some models used in hydrogeology contain millions of cells [*Mariethoz et al.*, 2009]. In petroleum engineering, the problem is even more acute since high-resolution models are used to simulate complex phenomena of multiphase, density-driven flow. The approach often adopted is then to calibrate (optimize) one realization at a time using optimization techniques.

[5] Therefore, depending on the computational burden involved, it may be appropriate to perform either Bayesian inversion (McMC) or optimization of one realization at a time (less CPU demanding but not consistent with Bayes' rule). The framework we present in this paper (iterative spatial resampling, ISR) allows dealing with both Bayesian inversion and optimization aspects. It is emphasized that our method is applicable in conjunction with any conditional geostatistical simulation method, whether it relies on hypotheses of multi-Gaussianity or not, and whether it generates continuous or categorical variables. In addition, we present a stopping criterion for optimizations, inspired from importance sampling, which allows approximating the posterior distribution at a lesser cost.

[6] This paper is organized as follows. Section 2 introduces the concept of perturbation by ISR, explores its properties for both Bayesian inversion and optimization, and performs numerical tests. Section 3 applies the method on a synthetic heterogeneous channelized aquifer to evaluate the posterior distribution using both Bayesian and optimization approaches.

## 2. Methodology

### 2.1. Bayesian Framework

[7] Formulated in Bayesian terms, the hydrogeological inverse problem consists of obtaining samples from a posterior distribution of models $f(\mathbf{m}|\mathbf{d})$ conditioned to a set of observed state data $\mathbf{d}$:

$$f(\mathbf{m}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{m})f(\mathbf{m})}{f(\mathbf{d})}. \quad (1)$$

In that formulation, the prior distribution $f(\mathbf{m})$ can be sampled by performing stochastic realizations not conditioned to the state variables $\mathbf{d}$. The likelihood function $L(\mathbf{m}) = f(\mathbf{d}|\mathbf{m})$ defines the probability of observing the actual measured state variables $\mathbf{d}$ (the data) given a certain model $\mathbf{m}$. It is a measure of how good the model $\mathbf{m}$ is in fitting the data. Computing the likelihood of a model $L(\mathbf{m})$ generally requires running a forward problem, denoted $\mathbf{d} = g(\mathbf{m})$. Choosing a particular likelihood function essentially amounts to deciding what is meant by "good-enough fit". It is a modeling decision that can be based on the distribution of measurement errors (which can be known for certain measurement devices) or can be subjectively taken. The very existence of the posterior relies on a likelihood function being defined, and this is the prerequisite of any Bayesian inversion. Hence, all methods presented in this paper assume that the likelihood function is given. Note that all optimization methods (Bayesian or not) need to define what a "good-enough fit" is, either under the form of a likelihood function or by choosing some kind of stopping criterion for search algorithms.

[8] *Tarantola* [2005] gives a comprehensive overview of the available exact methods to obtain samples representative of $f(\mathbf{m}|\mathbf{d})$. Among them, rejection sampling [*von Neumann*, 1951] and Metropolis sampling [*Metropolis et al.*, 1953] are often used. None of these methods requires the definition of the density $f(\mathbf{d})$. Rejection sampling is based on the fact that $f(\mathbf{m}|\mathbf{d})$ is a subset of $f(\mathbf{m})$, and therefore it can be evaluated by subsampling the prior. The approach consists in generating candidate models $\mathbf{m}^*$ that are samples of $f(\mathbf{m})$ and to accept each of them with a probability:

$$P(\mathbf{m}^*) = \frac{L(\mathbf{m}^*)}{L(\mathbf{m})_{\max}}, \quad (2)$$

where $L(\mathbf{m})_{\max}$ denotes the supremum, which can be any number equal to or above the highest likelihood value that can be taken by $L(\mathbf{m})$. Note that a higher supremum does not affect the accuracy of the sampling, but it can dramatically affect its performance. The distribution of the resulting samples follows $f(\mathbf{m}|\mathbf{d})$. Since it requires a large number of evaluations of $g(\mathbf{m})$, the rejection method is inefficient, but it will serve as a reference sampler in this paper.

[9] The Metropolis algorithm [*Metropolis et al.*, 1953] is able to perform a reasonably equivalent sampling by forming a Markov chain of models, such that the steady-state distribution of the chain is precisely the posterior distribution that one wishes to sample from. It is similar to a random walk that would preferentially visit the areas where $f(\mathbf{m}|\mathbf{d})$ is high. One issue with Metropolis samplers is that it is difficult to assess whether mixing of the chain (convergence) occurred. In addition, to ensure uniform sampling, each sample should come from a different Markov chain, and each independent chain should be carried on until a burn-in period is over. Since this requirement dramatically increases the cost of each sample, *Tarantola* [2005] suggests keeping only 1 every $m$ samples, where $m$ should be large enough for the chain to "forget" the previously accepted models.

[10] In this paper, we use a version of the Metropolis algorithm proposed by *Mosegaard and Tarantola* [1995]. To apply it, one needs to design a random walk that samples the prior. At each step $i$, it moves according to the following rules:

[11] (1) If $L(\mathbf{m}^*) \geq L(\mathbf{m}_i)$, move from $\mathbf{m}_i$ to $\mathbf{m}^*$.

[12] (2) If $L(\mathbf{m}^*) < L(\mathbf{m}_i)$, randomly choose to move to $\mathbf{m}^*$ or stay at $\mathbf{m}_i$, with the probability $L(\mathbf{m}^*)/L(\mathbf{m}_i)$ of moving to $\mathbf{m}^*$.

[13] The movement (or transition) from a model $\mathbf{m}_i$ to a model $\mathbf{m}_{i+1}$ is accomplished by drawing a candidate model $\mathbf{m}^*$ from the proposal distribution $Q(\mathbf{m}^*|\mathbf{m}_i)$, which denotes the probability density function of the transition from the model $\mathbf{m}_i$ to the model $\mathbf{m}^*$. The method requires that the proposal density is symmetric (or reversible), such that $Q(\mathbf{m}_i|\mathbf{m}^*) = Q(\mathbf{m}^*|\mathbf{m}_i)$.

[14] Previous studies have investigated Markov chains applied to spatially dependent variables, using different proposal (or perturbation) mechanisms. *Oliver et al.* [1997] create an McMC by updating one grid node of a geostatistical realization at each step. The method is very inefficient because it asks for a forward problem run after updating each node, which is not feasible for real-world grids. *Fu and Gomez-Hernandez* [2008] dramatically

accelerate the method by updating many grid nodes at the same time. They introduce the blocking Markov chain Monte Carlo (BMcMC) method that incurs local perturbations by successively resimulating a square area of the realizations (a block). The BMcMC method has been used for sampling the posterior distribution of synthetic inverse problems in a multi-Gaussian framework [*Fu and Gomez-Hernandez*, 2009].

[15] Optimization methods aim at finding realizations that maximize the likelihood $f(\mathbf{d}|\mathbf{m})$. They do not allow characterizing $f(\mathbf{m}|\mathbf{d})$ but are often used since they are much more efficient than sampling algorithms. These methods repeatedly update an initial solution to minimize an objective function, often measuring a misfit to measured data. Although regularization terms can be added to make the perturbed models look more realistic, prior constraints are often minimal. One can use either gradient-based [*Carrera and Neuman*, 1986] or gradient-free [*Karpouzos et al.*, 2001] methods. Since they search in a stochastic manner, gradient-free methods are less prone to be trapped in local minima (i.e., it is guaranteed that the global minimum is found after an infinite number of iterations). Upon convergence, a single calibrated solution is obtained. When several local minima are present, one can obtain alternative solutions by repeating the optimization procedure using different starting points.

[16] Among the optimization techniques, simulated annealing [*Kirkpatrick et al.*, 1983] has been extensively used to solve inverse groundwater modeling problems [e.g., *Pan and Wu*, 1998; *Zheng and Wang*, 1996]. Genetic algorithms [*Fraser*, 1957; *Goldberg*, 1989] have been used for identifying structures in hydraulic conductivity fields [*Karpouzos et al.*, 2001]. *Alcolea and Renard* [2010] apply the BMcMC method and simulated annealing to optimize non–multi-Gaussian random fields generated using the *impala* multiple-point simulation code (J. Straubhaar et al. An improved parallel multiple-point algorithm, submitted to *Mathematical Geosciences*, 2010).

[17] The gradual deformation method (GDM) [*Hu*, 2000; *Hu et al.*, 2001] and the probability perturbation method (PPM) [*Caers*, 2003; *Caers and Hoffman*, 2006; *Johansen et al.*, 2007] proceed by combining uniformly sampled realizations. By adjusting a single parameter, they allow obtaining a smooth transition from one simulation to another while preserving a prior structural model. Therefore, finding a calibrated realization can be accomplished by a series of 1D optimizations. These methods have been successfully applied in hydrogeology and petroleum engineering [*Le Ravalec-Dupin and Hu*, 2007; *Le Ravalec-Dupin*, 2010; *Llopis-Albert and Cabrera*, 2009; *Ronayne et al.*, 2008].

[18] In this paper we present a transition kernel (iterative spatial resampling, ISR) that can be used either as a proposal distribution with McMC sampling methods or as a perturbation strategy when optimization is used. We show that, in both cases, it yields accurate sampling of the posterior. To validate the results, we use the rejection sampler as a reference.

### 2.2. Reproduction of the Prior

[19] A premise of both rejection and Metropolis samplers is that proposal models $\mathbf{m}^*$ have a nonzero prior probability. It is not possible for a rejection sampler to produce samples with zero prior probability. We will define as a bounded prior a prior probability density for which there exist models $\mathbf{m}$ that have zero probability density. For example, in a Boolean model where simulated objects all exhibit the same (deterministic) direction, any model with different object directions has zero probability of occurrence. On the other hand, examples of unbounded prior densities are the multi-Gaussian model and the Markov random field model [*Besag and Kopperberg*, 1995; *Tjelmeland and Besag*, 1998]. With unbounded priors, all models are possible, even those with extremely low prior density. For example, in the standard multi-Gaussian, the model $\mathbf{m} = 0$ (a model with zeros everywhere and hence variance = 0) is highly improbable but not impossible. Actually, this very small probability can be regarded as a mathematical artifact because the chance of sampling it is negligible.

[20] Application of Bayes' rule (equation (1)) assumes that an intersection between prior and likelihood exists. In other terms, if all samples that have a (practically) nonzero prior probability also have a (practically) zero likelihood, this means that prior and likelihood are incompatible and the solution is nonidentifiable [*Carrera and Neuman*, 1986]. The posterior is then undetermined since it results, after normalization, in a division of zero by zero. In such cases, prior and likelihood do not intersect, Bayes' rule is not applicable, and a modeling decision has to be taken whether to put in question the data, the likelihood model, or the prior. This decision should be motivated by reconsidering the basis adopted to define the prior as well as the confidence given to the data.

[21] In the absence of intersection between prior and likelihood, data-driven inverse modeling techniques match the data at the price of producing models with zero or very low prior probability (for example, not preserving the specified spatial dependence). On the other hand, prior-driven techniques favor the prior and hence the statistics reflected in prior models (such as a variogram), but may be unable to achieve good fits to conflicting data. By opting for a certain inverse modeling technique, the modeler decides which piece of information should prevail in case of incompatibility.

[22] Among data-driven methods, we mention the gradual deformation method (GDM) [*Caers*, 2007; *Hu and Le Ravalec-Dupin*, 2004; *Le Ravalec-Dupin and Noetinger*, 2002; *Liu and Oliver*, 2004], the quasi-linear method [*Kitanidis*, 1995; *Zanini and Kitanidis*, 2009], and the regularized pilot points method (RPPM) that considers constraints imposed by both data and prior and adjusts these constraints using a weighted regularization term [*Alcolea et al.*, 2006; *Doherty*, 2003; *Hendricks-Franssen et al.*, 2004]. In the class of prior-driven methods, one can find the probability perturbation method [*Caers and Hoffman*, 2006] and the blocking moving window [*Alcolea and Renard*, 2010]. ISR, the technique we present in this paper, also belongs to this class of prior-driven methods.

### 2.3. Iterative Spatial Resampling

[23] Let a model $\mathbf{m}_i = \{Z_i(\mathbf{x}_1), \ldots, Z_i(\mathbf{x}_M)\}$ be a realization of a random variable $Z$ discretized on a grid with $M$ nodes. Unconditional realizations of $Z$ are considered samples of the prior, whether this prior is explicitly stated such as is the case of a multi-Gaussian model or whether this prior is defined by a given stochastic algorithm with a set of parameters

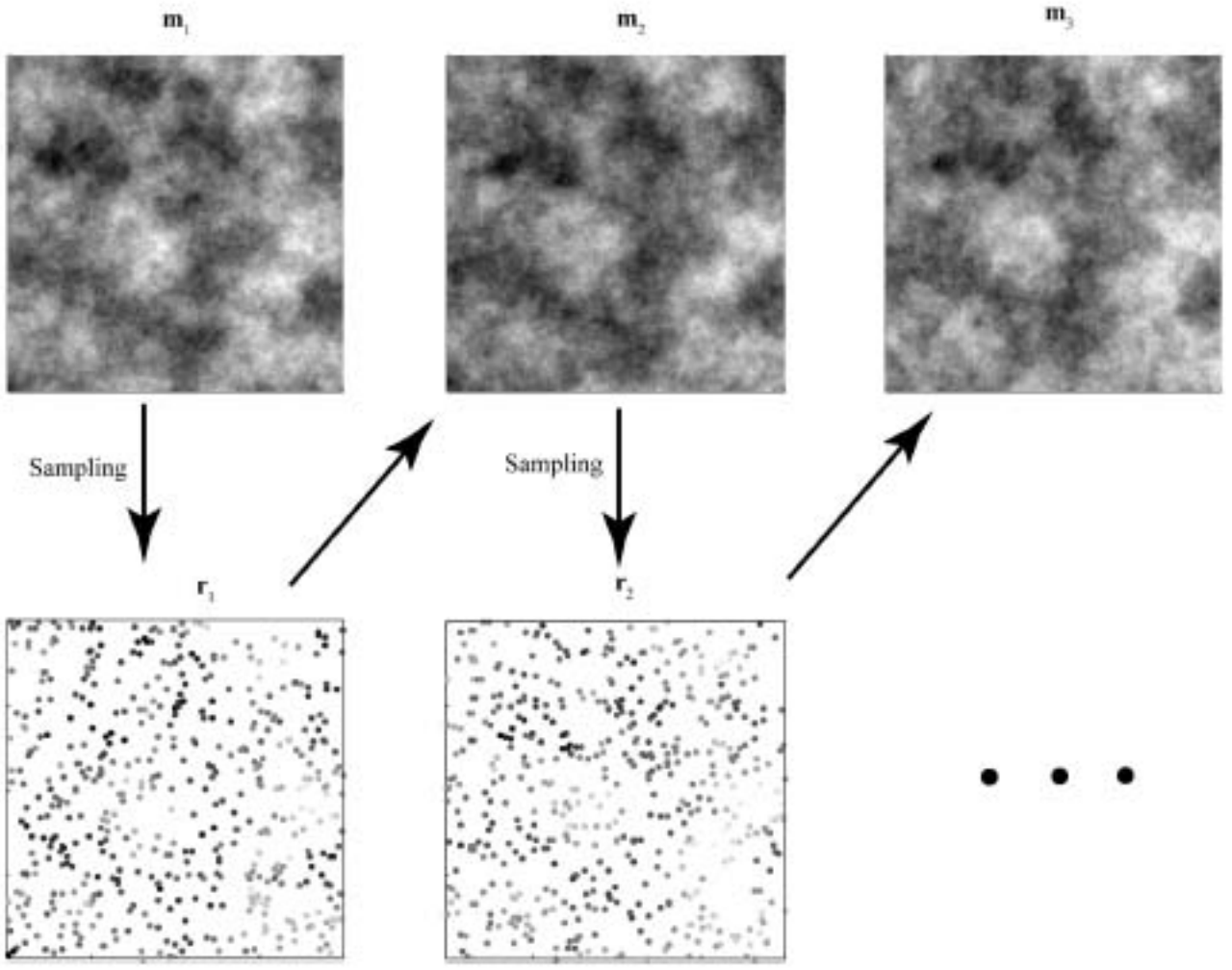


**Figure 1.** Sketch of the ISR method. An initial realization $\mathbf{m}_1$ is sampled randomly to obtain the subset $\mathbf{r}_1$, which is used as conditioning data for generating another realization $\mathbf{m}_2$. $\mathbf{m}_2$ displays local features similar to those of $\mathbf{m}_1$ due to the constraints imposed by the conditioning data, but it is also different since the simulation has produced new values at nonconditioned locations.

(A. Boucher, Stanford Center for Reservoir Forecasting, unpublished, 2007).

[24] To implement sampling and searching strategies, one needs to create a chain of dependant realizations. Consequently, one wants to draw proposal models $\mathbf{m}^*$ not from $f(\mathbf{m})$ but from $Q(\mathbf{m}|\mathbf{m}_t)$, where $\mathbf{m}_t$ is the previous model in the chain. To preserve the spatial continuity defined by the geostatistical simulation algorithm, the conditional term should ideally be incorporated in the method used to generate the realizations. Since most simulation methods also allow generating realizations conditioned to points data, we propose to use this conditioning capability to impose a conditional term on the prior. More specifically, dependence between $\mathbf{m}^*$ and $\mathbf{m}_t$ is introduced by extracting a subset of realization $\mathbf{m}_t$ as $n$ randomly located points $\mathbf{r}_t = \{Z_t(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$ and to impose these points as conditioning data to generate $\mathbf{m}^*$. The amount $n$ is a tuning parameter. Proposal models are drawn from $f(\mathbf{m})$, but at the same time they depend on $\mathbf{r}_t$, itself a subset of $\mathbf{m}_t$.

[25] Creating a Markov chain using ISR is accomplished by performing the following steps:

[26] 1. Generate an initial model $\mathbf{m}_1$ using a geostatistical simulation algorithm, and evaluate its likelihood $L(\mathbf{m}_1)$.

[27] 2. Iterate on $t$.

[28] a. Select randomly a subset $\mathbf{r}_t = \{Z_t(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$ of $n$ points belonging to $\mathbf{m}_t$.

[29] b. Generate a proposal realization $\mathbf{m}^*$ by conditional simulation using $\mathbf{r}_t$ and the same geostatistical model with a new random seed.

[30] c. Evaluate $L(\mathbf{m}^*)$.

[31] d. Accept or reject $\mathbf{m}^*$. If accepted, set $\mathbf{m}_{t+1} = \mathbf{m}^*$, otherwise *go back* to a (i.e., do not increment $t$). If the acceptance criterion is the one proposed by *Mosegaard and Tarantola* [1995], the chain is a Metropolis sampler.

[32] The method is illustrated in Figure 1, where an initial sequential Gaussian simulation (SGS) realization is iteratively perturbed. However, any simulation method can be used, as long as it is able to produce conditional simulations. Using ISR when actual conditioning points data are present (for example, corresponding to field measurements) can be accomplished in a straightforward manner by adding, at each iteration, the "real" conditioning data to the sampled set $\mathbf{r}$.

[33] The total amount of conditioning data retained, namely $n$, allows determining the strength of the dependency between two successive members of the chain $\mathbf{m}_t$ and $\mathbf{m}_{t+1}$. Note that this amount can be conveniently defined as a fraction $\varphi$ of the total number of nodes in $\mathbf{m}_t$.

[34] It is important to note that the selected data set $\mathbf{r}$ follows by construction the same spatial continuity as imposed by the geostatistical algorithm; hence the resulting perturbed realization will, by construction, have the same spatial continuity as the initial realization. Both have a nonzero prior probability. The only requirement is that the conditioning on $\mathbf{r}$ is correct or, in other words, that the conditioning method does not introduce artifacts into the simulation, nor does it artificially affect uncertainty in the neighborhood of the points data.

[35] If $f(\mathbf{m})$ is a nonstationary model (for example, containing a trend), the method applies equally well because uniformly sampling a nonstationary realization results in a nonstationary set of sampled points $\mathbf{r}$. It is obvious that the method works for both categorical and continuous variables.

[36] Note that methods previously used in the context of McMC [*Alcolea and Renard*, 2010; *Oliver et al.*, 1997] also rely on the use of conditioning points data, but they are focused on local perturbations between the realizations in the Markov chain. The main difference is that they update one grid node or local group of nodes at one step, and then update another area at the next step. Because the updated area is often different at each step, the search pattern in the solution space at step $i$ tends to be orthogonal to the search direction at step $i - 1$ (in fact, it is not orthogonal when updated areas overlap, but this is not often the case). When high-dimensional spaces are explored, searching in orthogonal directions can be inefficient. In addition, with some simulation methods, and depending on the conditioning technique, resimulating local areas is prone to create artifacts in the simulations.

### 2.4. Sampling Properties of ISR

[37] Several factors may affect the accuracy of the Metropolis sampler. We mentioned above that each sample should be obtained from a different, independent chain. Obtaining them from a single Markov chain, even if the samples are far apart in the chain, is an approximation. Moreover, convergence of the chain must be reached before performing any sampling, and this is difficult to assess.

[38] In theory, the proposal distribution $Q(\mathbf{m}_t|\mathbf{m}^*)$ is symmetric. In Figure 1, consider the set of points $\mathbf{r}_1$. If one would use it as conditioning data for a new realization, all possible outcomes would have an equal likelihood of being drawn as long as the conditional simulation samples uniformly. Therefore, the outcome has an identical probability of being $\mathbf{m}_1$, $\mathbf{m}_2$, or a set of other possible models. However, since geostatistical simulations are algorithmically defined (A. Boucher, unpublished, 2007), they may not offer perfect conditioning, thus making the proposal distribution possibly nonsymmetric. For example, conditioning with kriging in the multi-Gaussian case is a very accurate conditioning method, but it is not the case for SGS with a limited neighborhood [*Emery*, 2004].

[39] We set up a simple synthetic example to illustrate the properties of ISR within the Metropolis algorithm. The variable of interest $Z$ presents a multi-Gaussian exponential covariance model with an isotropic range of 10 grid nodes, a mean of 0, and a variance of 1. These characteristics constitute the prior distribution $f(\mathbf{m})$. The grid size is 50 by 50 nodes, and SGS [*Remy et al.*, 2009] is used to generate the realizations. Four numerical experiments are performed. These essentially aim at comparing the results of the Metropolis sampler described in the previous section with rejection sampling, which is known to be accurate. The numerical experiments consist of characterizing the prior by (1) unconditional sampling and (2) McMC sampling and characterizing the posterior by (3) rejection and (4) Metropolis sampling. These numerical experiments are described below.

[40] 1. We characterize the prior numerically by generating an ensemble $\mathbf{M}_r$ of 20,000 unconditional realizations, uniformly sampled from $f(\mathbf{m})$, and observe the variance of the simulated values (Figure 2a, black dashed line). The variance ranges approximately between 0.7 and 1.3. The variation in the realizations variance is due to statistical fluctuations.

[41] 2. We use an alternative way of characterizing the prior that uses the proposal density. It consists in performing a random walk using ISR where proposal models are systematically accepted (i.e., $f(\mathbf{d}|\mathbf{m}^*) = f(\mathbf{d}|\mathbf{m}_i)$, $\forall i$). In fact, it is a Metropolis algorithm that ignores the likelihood. If the requirements for a Metropolis algorithm are fulfilled, the steady state of such a chain should yield samples of the prior (a method suggested by *Mosegaard and Tarantola* [1995]). We generate another ensemble of 20,000 realizations $\mathbf{M}_m$ using such a Markov chain, with a fraction of resampled nodes of $\varphi = 0.1$ (i.e., $n = 250$ sampled nodes at each iteration) and keeping only one every $m = 100$ accepted models. Note that $m = 100$ is a large value that was chosen to have conditions close to an ideal sampler, but coming at a high CPU cost. The variance of the realizations in $\mathbf{M}_m$ and $\mathbf{M}_r$ have very similar distributions (Figure 2a, red and black dashed lines), showing that ISR did not induce significant deviations in the sampling.

[42] In the next numerical experiments, we use a likelihood that contains contradictory information with the prior (i.e., we purposely forge a case where likelihood and posterior do not intersect). We show that our sampling method is prior driven and therefore cannot create samples with (practically) zero prior probability. The aim in these experiments is to match a uniform reference field where $Z = 0$ on the entire domain. To this end, we define a likelihood function that is maximal when the variance of all $Z$ values is 0. It is expressed as

$$L(\mathbf{m}) = \exp\left(-\frac{\mathrm{var}(\mathbf{m})}{2\sigma^2}\right). \tag{3}$$

We set $\sigma = 0.2$ so that the likelihood quickly decreases with larger variances and approaches 0 when the variance equals the prior value of 1. This constraint of minimal variance is not compatible with the prior that imposes a unit variance, although fluctuations are possible within a certain range. ISR, as a prior-driven method, should be unable to produce samples with zero prior probability (i.e., not represented in $\mathbf{M}_r$). Note that data-driven methods could match such a constraint because $Z = 0$ is part of the Gaussian prior (it is an unbounded prior), but is very unlikely.

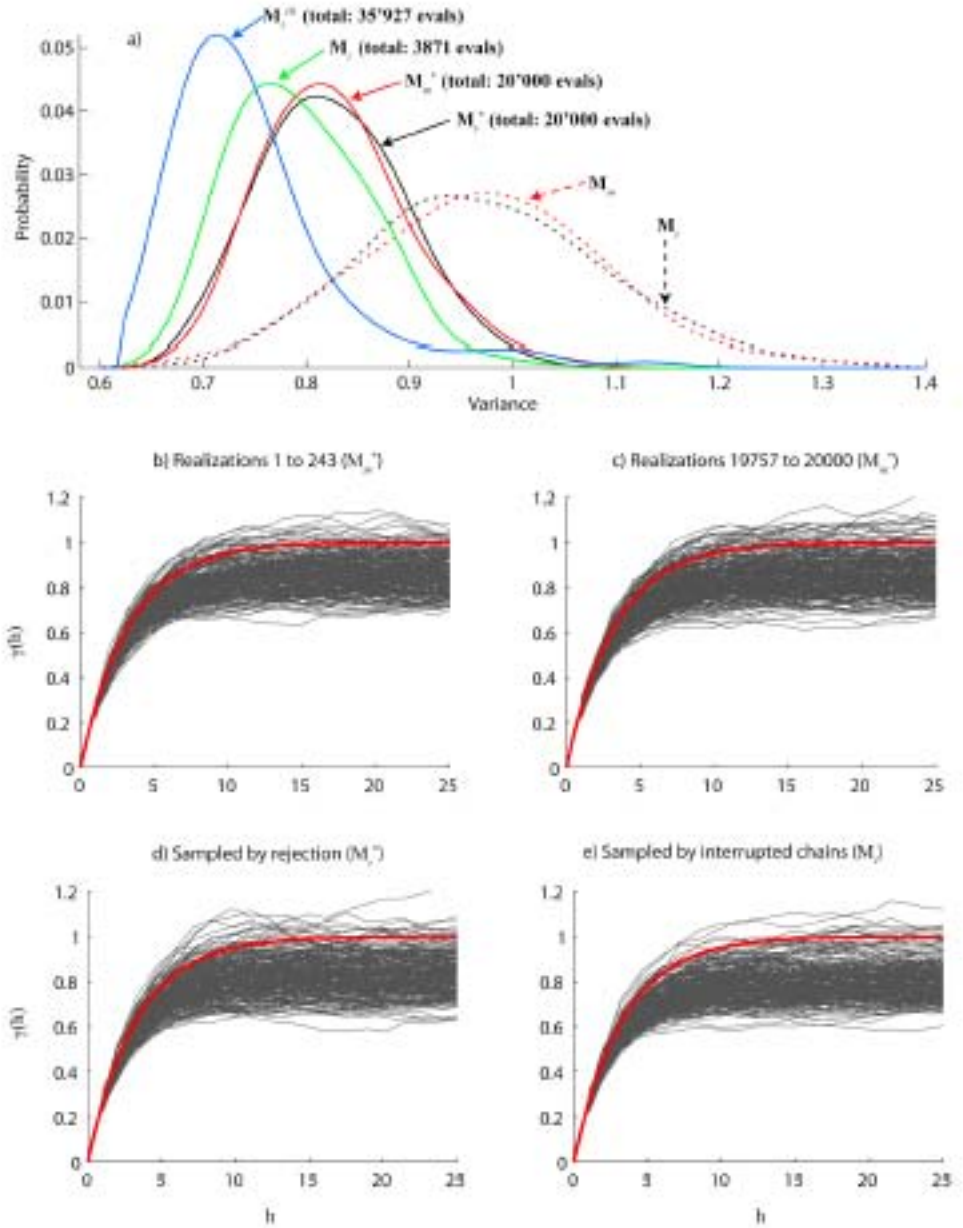


**Figure 2.** Comparison of different sampling methods with ISR. (a) Variance of realizations under systematic acceptance (dashed lines) and under minimum variance constraint (solid lines). The histograms are based on 243 samples for each ensemble. (b) Variograms of the first realizations of the chain. (c)Variograms of the last realizations of the chain. (d) Variograms of the realizations sampled by rejection. (e) Variograms of the realizations sampled by interrupted Markov chains.

[43] 3. Similarly to experiment 1, we want to accurately characterize the posterior distribution of the problem by rejection sampling. We apply rejection sampling to the realizations of $\mathbf{M}_r$, with the supremum $L(\mathbf{m})_{max}$ chosen such that it corresponds to a variance of 0.6, and 243 samples are accepted that constitute the reference posterior ensemble $\mathbf{M}_r^*$, representative of $f(\mathbf{m}|d)$ (Figure 2a, solid black line).

[44] 4. For the last numerical experiment, we perform a chain with the acceptance criterion of Mosegaard and Tarantola and the likelihood equation (3). Ideally, it should converge to $\mathbf{M}_r^*$. After a burn-in period ensuring that steady state of the Markov chain is reached, we perform iterations until an ensemble $\mathbf{M}_m^*$ of 20,000 samples is obtained, and we keep only one every $m = 100$ accepted models (Figure 2a, red solid line). Figure 3 shows that the number of iterations is large enough to ensure convergence. The variance distributions of $\mathbf{M}_r^*$ and $\mathbf{M}_m^*$ are similar, showing that the posterior is accurately sampled. To make sure that the prior spatial model is preserved, we compare in Figures 2b and 2c the experimental variograms of the realizations at the beginning and end of the Markov chain. Variogram shapes and ranges are preserved throughout the iterations, and therefore the Markov chain did not drift toward models with a (practically) zero prior probability. Note that the variograms present lower sills (smaller variances) than the variogram model used as input for SGS (red

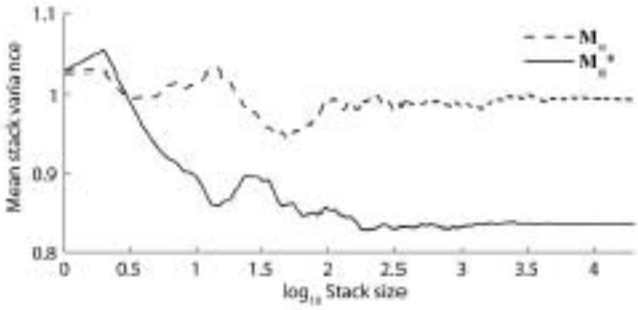


**Figure 3.** Convergence of the Metropolis chains used to obtain $\mathbf{M}_m$ and $\mathbf{M}_m^*$.

line). Indeed, the posterior models are not the same as the prior models since the inclusion of data yields a different ensemble. This is coherent with Bayes' rule that acknowledges the influence of the likelihood. In this case, prior models with a low variance have been sampled more often, but a zero variance cannot be reached since ISR is prior driven. For comparison, Figure 2d shows the experimental variograms of the models sampled by rejection ($\mathbf{M}_m^\sharp$), which are similar to the ones sampled by Metropolis ($\mathbf{M}_m^*$). These four numerical experiments show that, in this case, both rejection method and Metropolis sampling with ISR give similar results.

### 2.5. Using ISR for Optimization

[45] *Mosegaard and Tarantola* [1995] indicate that their sampling method can also be used for optimization. To this end, one can create a chain of ever-improving realizations using for acceptance criterion (step 2d of the ISR algorithm as described in section 2.3):

$$\text{if } L(\mathbf{m}^*) \geq L(\mathbf{m}_i), \text{ accept } \mathbf{m}^*. \tag{4}$$

[46] The resulting McMC process is a stochastic search for a single calibrated model. The search strategy of ISR performs by successive steps in random directions and of random size (step size is random, but its distribution is controlled by $\varphi$). When large steps occur, it allows exploring various regions of the solution space. Large steps are also an opportunity to leap out of local minima. On the other hand, the occurrence of small steps allows fine tuning suboptimal solutions. Since the search is stochastic, the global minimum will eventually be reached after an infinite number of iterations. However, in most practical applications, it will remain in a local minimum (i.e., suboptimal). Figure 4 schematically depicts how a local minimum is searched in a simple 2D solution space. The background image represents the real, unknown solution space, with a local minimum in the center of the image and the global minimum in the top right. In this case, the search remains in the local minimum because criterion (4) is used and the number of iterations is finite. Since (4) only considers the rank of a proposal solution compared to a previous one, the search is similar to the minimization of an objective function. In this sense optimization with ISR is similar to evolutionary strategies where probabilistically generated individuals are sequentially improved while only the best one is preserved [*Bayer and Finkel*, 2004]. The likelihood of

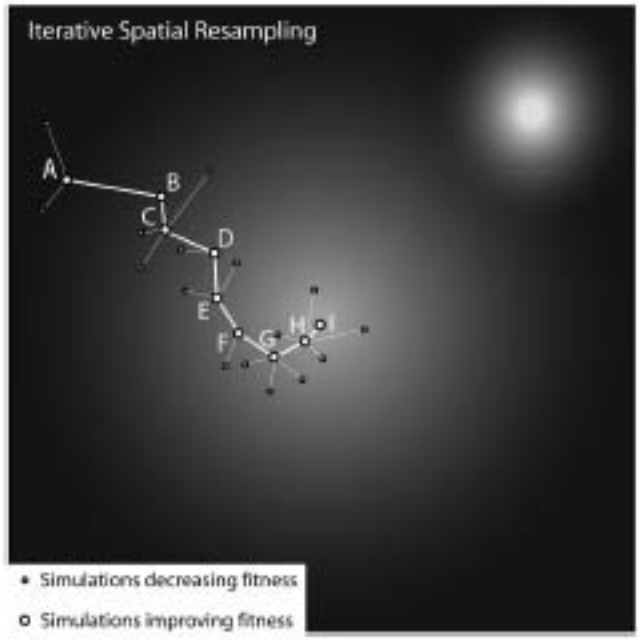


**Figure 4.** Schematic representation of the search strategy used by ISR in a simple solution space having two degrees of freedom. The background color represents the actual unknown solution space, with values ranging from black (bad solution) to white (good solution). A local minimum lies in the center of the domain, while the global minimum is in the top right.

the final solution depends on algorithmic parameters such as the stopping criterion used.

[47] So far a constant fraction of resampled nodes has been considered, but alternatives can be envisioned for optimization. For example, $\varphi$ can increase at each iteration $i$. The optimization starts with large steps (i.e., the solution is largely perturbed at the exploration phase) and finishes

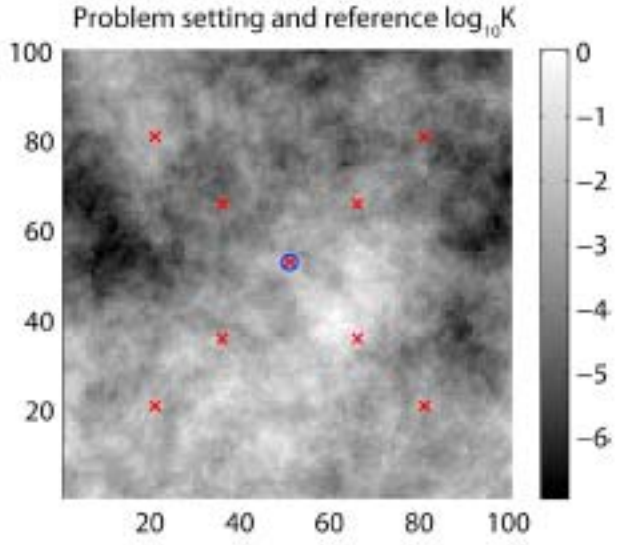


**Figure 5.** Reference logarithm of hydraulic conductivity field in meters per second. The blue circle depicts the location of the pumping well, and the red crosses indicate the locations of the observation wells.

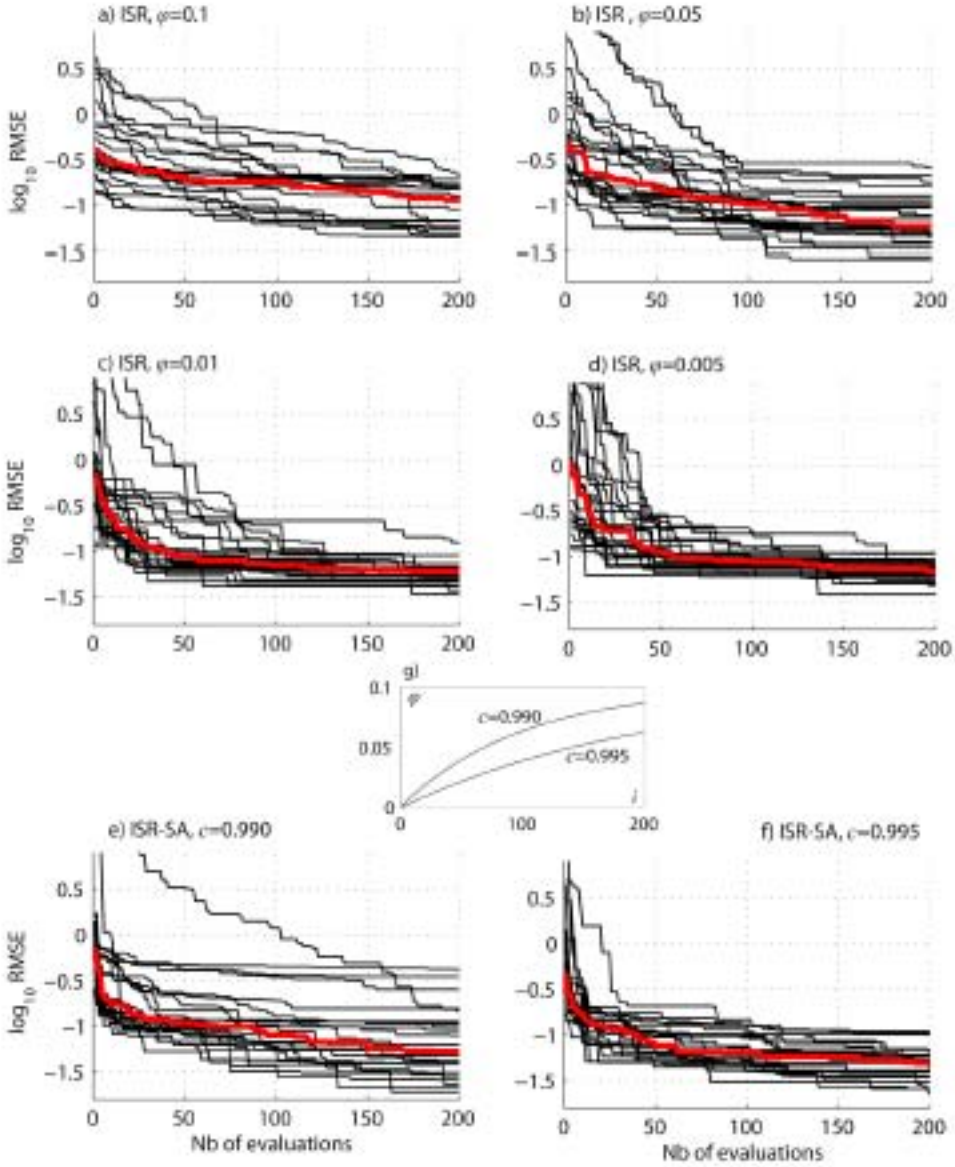


**Figure 6.** Optimization performance assessment for six different values of φ. The evolution of each optimization is marked by a thin black line, and the median evolution computed from 25 optimizations is depicted by a red bold line.

with small steps (narrowing phase). One possible way to accomplish this is with a power law:

$$\varphi(i) = \varphi_{\max}(1 - c^i), \quad 0 \le c \le 1. \tag{5}$$

[48] The larger $c$ is, the slower $\varphi(i)$ will reach $\varphi_{\max}$. This is similar to simulated annealing, with the parameter $c$ defining the cooling schedule and $\varphi_{\max}$ is the maximum value of $\varphi$ (after an infinite number of iterations). Yet, alike other simulated annealing algorithms, adjusting $c$ can be tedious. Nevertheless, we will see in section 2.6 that using (5) can accelerate the convergence compared to keeping $\varphi$ constant. Note that a varying $\varphi$ cannot be used in the context of a Bayesian inverse problem (Metropolis acceptance criterion) because it would make the proposal density non-symmetric. In the context of optimization, (5) is applicable since the problem is not to sample the posterior, but to reach a local minima quickly.

### 2.6. Sensitivity to φ

[49] $\varphi$ is the only parameter required when ISR is used for optimization. In order to evaluate its sensitivity on the optimization convergence speed, we set a simple flow problem and perform several optimizations with different values of $\varphi$.

**Table 1.** RMSE of the Different Optimization Runs After 200 Iterations

| | Minimum RMSE | Median RMSE | Maximum RMSE |
|---|---|---|---|
| ISR, $\varphi$ = 0.1 | 0.0451 | 0.1141 | 0.2124 |
| ISR, $\varphi$ = 0.05 | 0.0271 | 0.0591 | 0.2610 |
| ISR, $\varphi$ = 0.01 | 0.0340 | 0.0607 | 0.1242 |
| ISR, $\varphi$ = 0.005 | 0.0384 | 0.0685 | 0.1096 |
| ISR-SA, $c$ = 0.990 | 0.0190 | 0.0529 | 0.4094 |
| ISR-SA, $c$ = 0.995 | 0.0231 | 0.0481 | 0.1087 |

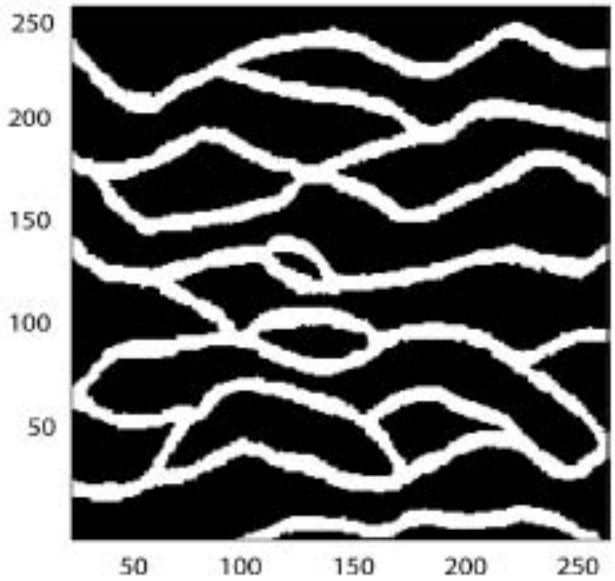


Figure 7. Training image used to model for the sand and clay spatial distribution.

[50] The problem setting consists of a square aquifer of 100 m by 100 m, discretized in 100 by 100 elements (Figure 5). A reference log $K$ field is generated using SGS. The spatial model for hydraulic conductivity is multi-Gaussian, with an isotropic exponential variogram of range 90 m. The mean logarithm of the hydraulic conductivity is −3, with a variance of 2. The upper and lower sides of the model are no-flow boundaries, a fixed head boundary condition of 1 m is set on the left side, and a fixed head boundary condition of 0 m is set on the right side. A pumping well extracting 0.003 $m^3/s$ is set in the center of the domain (blue circle), and nine observation wells are positioned at the locations of the red crosses in Figure 5 (the pumping well is also an observation well). The problem is solved in steady state. The nine head measurements are the only constraints used to solve the inverse problem, i.e., to find log $K$ fields that match these data. We do not impose any conditioning points data in order to not over-constrain the prior.

[51] Proposal solutions are generated using SGS with the true variogram model. The acceptance criterion is (4), and therefore the likelihood has the role of an objective function to minimize. This objective function measures the root-mean-square error (RMSE) between calculated and measured heads at the nine observation wells. With such loose constraints, we ensure that the problem is severely ill posed and that the solution space has multiple local minima.

[52] Six series of runs are performed. The first four series (ISR) use fixed $\varphi$ values of (a) 0.1, (b) 0.05, (c) 0.01, and (d) 0.005. The last two series (ISR-SA) use a varying sampled fraction $\varphi$ according to (5). For each series, 25 optimizations are performed with ISR; each optimization is carried out for $i$ = 200 iterations. Figure 6 displays the evolution of each optimization (thin black line) and the median of each series (bold red line). The parameters of the simulated annealing cooling schedule are (Figure 6e) $c$ = 0.990 and (Figure 6f) $c$ = 0.995, with $\varphi_{max}$ = 0.1 for both series. Figure 6g shows the evolution of $\varphi$ as a function of the iterations for both cooling schemes, with $c$ = 0.990 representing a fast cooling and $c$ = 0.995 reperesenting a slower cooling. Table 1 provides a summary of the RMSE values obtained with each series of runs. Median values illustrate the overall performance for each series and minimum/maximum values help identify the stability of the optimization behavior.

[53] Nearly all parameters are, on average, able to reduce the RMSE by more than 1 order of magnitude in 200 iterations. The only exception is $\varphi$ = 0.1, whose poor performance can be explained by too small steps between one solution and another (large fraction of resampled nodes). Although this does not prevent reaching a local minimum, it can significantly reduce the convergence rate if the topography of the objective function is very flat. All other $\varphi$ values tested achieve similar median fits. Constant resampling with $\varphi$ = 0.05 performs slightly better for the median and the best fit than smaller $\varphi$ values, but this comes at the price of having some optimizations that did not converge,

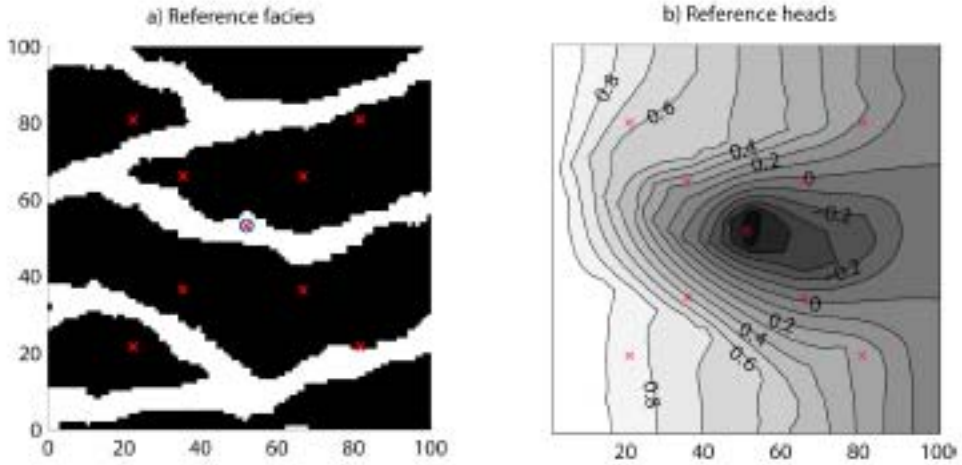


Figure 8. Reference field used for the synthetic test case. The blue circle marks the location of the pumping well, and the red crosses indicate observation wells. (a) Reference facies. (b) Corresponding reference heads.

Table 2. Test Case Using ISR With Direct Sampling[a]

| | Min RMSE | Median RMSE | Max RMSE | Number of Evaluations |
|---|---|---|---|---|
| Prior | 0.0435 | 0.9977 | 3.5924 | 100,000 |
| Rejection | 0.0435 | 0.0754 | 0.1155 | 100,000 |
| Metropolis | 0.0486 | 0.0873 | 0.1294 | 26,753 |
| Interrupted MC | 0.0305 | 0.0745 | 0.1098 | 8108 |

[a]Ranges of RMSE and number of forward problem evaluations for the ensembles sampled with different methods.

again due to the relatively small steps. This is shown by the maximum RMSE that is larger for $\varphi = 0.05$.

[54] The same phenomenon occurs with ISR-SA. If the large steps at the beginning of the optimization do not yield models close to a good fit, the smaller steps that occur later can only provide limited improvements. The search then remains away from good areas, and the corresponding solutions show poor fits, even after a large number of iterations. Conversely, if a suboptimal solution is reached in the initial phase of the optimization, the smaller steps that occur later allow fine adjustment. This dual and unstable behavior explains the presence of both the best and worst fits of all series when ISR-SA is used with a quick cooling schedule ($c = 0.990$). With a slower cooling ($c = 0.995$), it is less pronounced. Due to this high sensitivity to the initial sample, a possible strategy could be to choose the starting point of the optimization as the best of a small set of randomly generated realizations.

[55] ISR-SA has the potential of achieving better fits, but it provides only slightly better median convergence. The price to pay is a high sensitivity to the cooling parameter that may be difficult to adjust in practice. Conversely, ISR does not require the adjustment of cooling parameters, and we think this is a major advantage from a practical point of view. We tested a large array of $\varphi$ values, with a factor 20 between the lowest and highest $\varphi$ values. Corresponding RMSE values vary only with a factor 2; therefore, it seems that ISR is not very sensitive to the parameter $\varphi$, at least for the present case. This is fortunate because it eases the adjustment of optimization parameters.

### 2.7. Approximating the Posterior with Multiple Optimizations

[56] Consider $n$ independent Markov chains, each using acceptance criterion (4) to define the models that are accepted in the chain. Taking one optimized model per chain yields an ensemble of $n$ samples, all of them having a nonzero prior probability and fitting the data. However, Bayes' rule may not have been observed since models are sampled from a subset of the prior that may not reflect the exact posterior. Samples can belong to the posterior (in the sense that they match the data well), but they are not necessarily distributed according to the posterior. Therefore, a bias is introduced on the modeling of uncertainty (here we use the term "bias" in the sense of a faulty sampling design).

[57] Such a procedure is a form of importance sampling. The central idea of importance sampling is that certain areas of the prior distribution have more impact on the posterior than others. Hence it may be preferable to avoid proposing samples in regions of low fit (see *Smith* [1997] for a comprehensive review). Instead of uniformly sampling from $f(\mathbf{m})$, one wishes to sample models from a biased distribution $f^*(\mathbf{m})$ that excludes areas of low fit. As a result, sampling is not as imposed by Bayes' rule, but according to a biased posterior. Importance sampling techniques provide an approximate compensation for such bias by introducing a weighting term in the probability of acceptance of a model $\mathbf{m}$, weights being given by the ratio of the priors $f(\mathbf{m})/f^*(\mathbf{m})$.

[58] Since importance sampling can greatly accelerate the sampling process, its use in the context of the hydrogeological inverse problem is appealing. However, applying the bias correction in practical cases is problematic because the ratio of priors is difficult to define. Without bias correction, there is no guarantee that samples obtained by multiple optimizations are even approximately distributed according to Bayes' rule. The distribution of the sampled models is dependent on the stopping criterion of the optimization process. If the number of iterations is too large, all optimizations converge to the global minimum. In addition to wasting CPU time, it results in an uncertainty smaller than desired. If the number of iterations is too small, a very large portion of the prior is sampled, yielding too high uncertainty. In other words, deterministic stopping criteria give little control of whether the data are over- or underfitted. We illustrate this problem with the example of variance minimization (Figure 2). We perform 243 optimizations using acceptance criterion (4), and we set as a stopping criterion a total of $i_{max} = 10$ accepted models. The resulting ensemble $\mathbf{M}_i^{10}$ (Figure 2a, blue line) presents a distribution of variance much narrower than what is found with rejection sampling (Figure 2a, black solid line). Moreover, the large amount of forward model evaluations (35,927) represents a waste of computer resources.

[59] In the case of ISR using acceptance criterion (4), the models in the Markov chain are drawn from the biased prior $f^*(\mathbf{m}) = f(\mathbf{m}|\mathbf{m}_i)$, which is the ensemble of all realizations obtained by extracting a subset $\mathbf{r}_i$ from the previous member of the chain $\mathbf{m}_i$. At each iteration, $f^*(\mathbf{m})$ is more biased toward high fits but still remains a subset of the prior, as shown in section 2.3. Therefore, a first bias is that the likelihood of the models is too high.

Figure 9. Representation of the ensembles of models obtained with different methods (150 models per method). Each column represents the results of a sampling method (unconditional prior, rejection sampler, Metropolis sampler, interrupted Markov chains). Each row corresponds to a different representation of the ensemble (mean heads, standard deviation of heads, probability of channels, multidimensional scaling mapping).

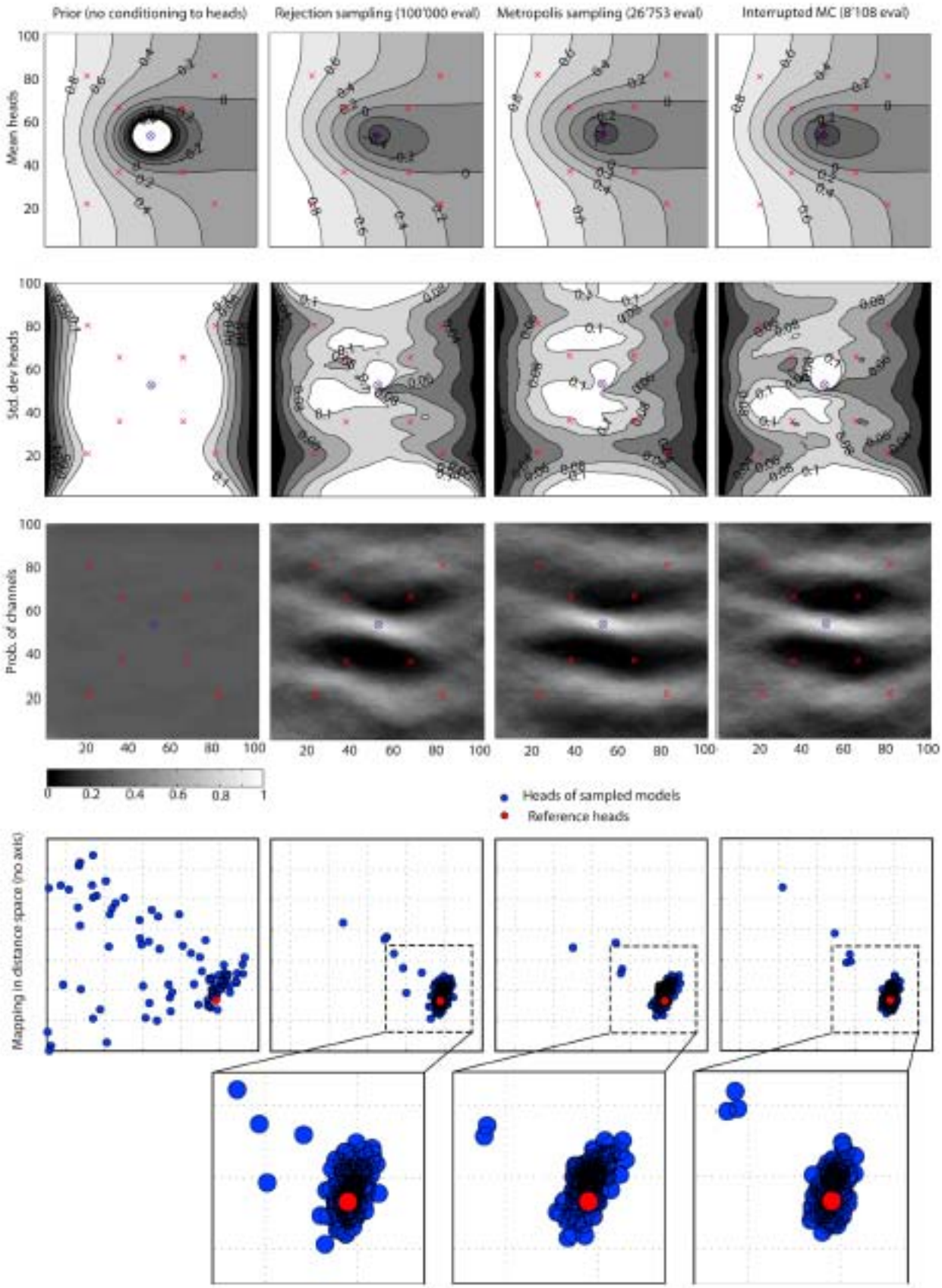
Prior (no conditioning to heads)
Rejection sampling (100'000 eval)
Metropolis sampling (26'753 eval)
Interrupted MC (8'108 eval)
Mean heads
Std. dev heads
Prob. of channels
Heads of sampled models
Reference heads
Mapping in distance space (no axis)

Figure 9

[60] As a practical bias correction, we propose to prematurely interrupt and sample the chain, with a criterion based on the likelihood. Our idea relies on the fact that for a proposal model $\mathbf{m}^*$ to be submitted to the possibility of acceptance, all previous models in the chain $\mathbf{m}_1, \ldots, \mathbf{m}_i$ must also have been submitted to this same possibility and rejected. In other words, the existence of a model is conditioned to the rejection of all of its predecessors. Hence, the probability that a model is even considered as a sample decreases with the iterations, which is a second bias on $f^*(\mathbf{m})$, but in the opposite direction. Models are increasingly likely to be accepted, but they are less and less likely to be submitted to the acceptance criterion. Although they are difficult to define precisely, both effects are opposite and may compensate each other.

[61] To obtain one sample by interrupted Markov chain, one needs to design an ever-improving Markov chain that accepts new members under condition (4). The chain should be interrupted following a stochastic stopping criterion similar to the acceptance rule of rejection sampling. This is can be accomplished in the following steps:

[62] 1. Define the supremum $L(\mathbf{m})_{max}$.

[63] 2. Generate an initial model $\mathbf{m}_1$ using a geostatistical simulation algorithm, and evaluate its likelihood $L(\mathbf{m}_1)$.

[64] 3. Iterate on $i$ until interruption:

[65] a. Select randomly a subset $\{\mathbf{r}_i = Z_i(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$ of $n$ points belonging to $\mathbf{m}_i$.

[66] b. Generate a proposal realization $\mathbf{m}^*$ by conditional simulation using $\mathbf{r}_i$ and the same geostatistical model with a new random seed.

[67] c. Evaluate $L(\mathbf{m}^*)$.

[68] d. Decide whether or not to interrupt the chain:

[69] i. Compute $P(\mathbf{m}^*) = L(\mathbf{m}^*)/L(\mathbf{m})_{max}$.

[70] ii. Draw $u$ in $U[0,1]$.

[71] iii. If $u \leq P(\mathbf{m}^*)$, accept $\mathbf{m}^*$ as a sample of the posterior distribution and interrupt the chain. If $u > P(\mathbf{m}^*)$, continue the chain.

[72] e. Attempt to narrow down $f^*(\mathbf{m})$: if $L(\mathbf{m}^*) \geq L(\mathbf{m}_i)$, set $\mathbf{m}_{i+1} = \mathbf{m}^*$, otherwise *go back* to a (i.e., do not increment $i$).

[73] The algorithm above is a form of rejection sampler that samples from the proposal models of a chain instead of uniformly sampling the prior. It is indeed a biased sampler compared to a rejection sampler (except for the first iteration, where it is exactly a rejection sampler). It is a heuristic way to quickly obtain an approximation of the posterior, and it does not replace exact samplers. However, it still accounts for the likelihood function, which is not the case with deterministic stopping criteria such as a fixed number of iterations, a maximum number of iterations without improvement, or a threshold in the objective function. More importantly, interrupting the chains reduces computational burden by skipping the unnecessary runs that incur overfitting. Incidentally, since each Markov chain is independent, the approach is straightforward to parallelize [*Mariethoz*, 2010].

[74] As a preliminary test of the interrupted Markov chains, we use it to evaluate the posterior distribution of the minimum variance problem (Figure 2). We generate 243 samples with interrupted Markov chains, using likelihood (3) and the same supremum value that was used for rejection sampling. The variance distribution of the resulting ensemble $\mathbf{M}_I$ is displayed in Figure 2a (green line), and the variograms of all models are shown in Figure 2e. As expected, $\mathbf{M}_I$ does not display exactly the same distribution of variance and variograms as the ensembles obtained with rejection and Metropolis samplers. However, the bias is less than with a fixed number of 10 iterations ($\mathbf{M}_I^{10}$), and it requires much less model evaluations. The algorithm of interrupted Markov chains is able to obtain a reasonable ensemble with only 3871 evaluations, whereas rejection and Metropolis sampling need 20,000 evaluations and a fixed number of 10 iterations requires 35,927 evaluations.

## 3. Test Case

### 3.1. Problem Setting

[75] One of the key features of ISR is that its principle is not associated with a specific simulation method or a certain type of spatial variability. In section 2, we presented ISR with multi-Gaussian examples. To demonstrate the general applicability of ISR, we define a new problem involving sand channels in a clay matrix and we use the direct sampling method (DS) to model it [*Mariethoz and Renard*, 2010; *Mariethoz et al.*, 2010]. This technique uses multiple-point statistics, which are well-suited to model a wide range of structural models, multi-Gaussian or not [*Caers*, 2003, 2005; *Guardiano and Srivastava*, 1993; *Hu and Chugunova*, 2008; *Journel and Zhang*, 2006; *Strebelle*, 2002]. DS has the particularity that it does not determine the values of the simulated nodes by drawing them from local probability distributions. Instead, values are directly sampled from a training image. Therefore, inverse modeling methods that rely on the perturbation of the local distribution, such as GDM and PPM, cannot be applied. Since DS allows conditioning to points data, we show that ISR can be applied.

[76] The spatial model of the sand/clay patterns is defined by the categorical training image displayed in Figure 7, representing sand channels in a clay matrix [*Strebelle*, 2002]. With this training image and the parameters described below, one realization is generated on a grid of 100 by 100 nodes, which is thereafter considered as the reference field (Figure 8a). Parameters of the simulation are a neighborhood of $n = 25$ nodes and a distance threshold set to $t = 0.04$. The meaning of these parameters is that, for any simulated node, the data event (pattern) made of the 25 closest neighbors is considered. Starting from a random location, the training image is scanned until encountering a node whose neighborhood matches at least 24 out of the 25 nodes searched for. (The parameter $t$ represents the fraction of mismatching nodes allowed, which here equals 1 since $0.04 \times 25 = 1$. Hence up to one mismatching node is allowed.) The value of this node is then assigned to the simulated location. The method reproduces the statistics of the training image up to the $n$th order [*Shannon*, 1948].

[77] Although multiple-point algorithms guarantee that conditioning data are locally honored, there may be artifacts in the neighborhood of the conditioning data [*Kjønsberg and Kolbjørnsen*, 2008]. Conversely, kriging offers perfect conditioning at the data locations and in the spatial relationships between conditioning data and surrounding locations. In the case of DS, when a data configuration observed in the simulation is not found in the training

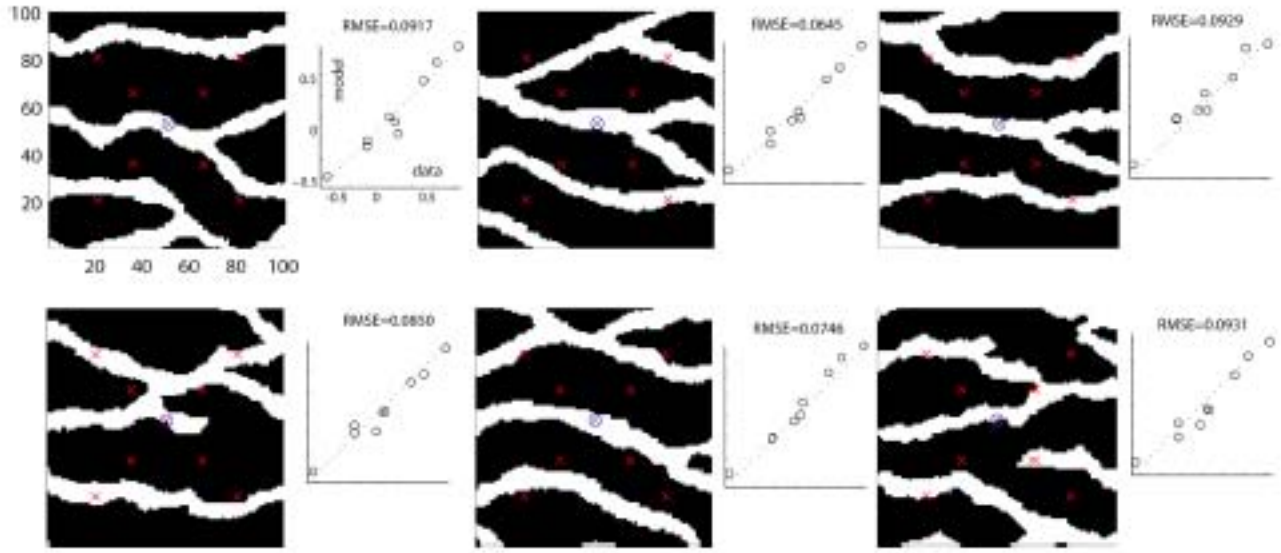


**Figure 10.** Six realizations (randomly chosen) out of the 150 sampled with the rejection method. The fits to data and the RMSE are shown on the right of each realization. The axes and labels are the same for all realizations, but they are only displayed for the top left image.

image, DS selects the best matching configuration of the training image. In such cases, patterns that are incompatible with the prior can occur in the simulation. If these patterns are in the neighborhood of data artifacts may appear, especially when large amounts of conditioning data are present, which is the case with ISR. Consistency among all patterns could be enforced using syn processing [*Mariethoz et al.*, 2010], which recursively unsimulates and resimulates nodes until all of them are compatible, but the method has a steep CPU cost. Instead, we use here a specific distance between data events (also described by *Mariethoz et al.* [2010]) that gives a larger relative weight to the nodes corresponding to data. However, perfect conditioning is not guaranteed.

[78] A uniform hydraulic conductivity value of $10^{-2}$ m/s is assigned to sand channels (Figure 7, white) and a value of $10^{-4}$ m/s to clays (Figure 7, black). The resulting hydraulic conductivity field is used in the same setting as the example in section 2.6 (Figure 5: one pumping well and nine observation wells, and the same boundary conditions). The resulting reference heads are displayed in Figure 8b. The head is known at the nine observation wells, and the RMSE of the calculated versus observed head is considered to evaluate a given solution.

### 3.2. Ensemble Solutions with Different Samplers

[79] The posterior distribution is characterized using different techniques. Table 2 provides the RMSE of the calculated heads and the number of forward problem evaluations for each sampling method. Figure 9 summarizes the results graphically. Each column represents a sampling method, and each row represents a different representation of the ensembles of models considered. The first row is the ensemble mean head, the second row is the head standard deviation, and the third row is the probability of occurrence of channels.

[80] In the fourth row, we use the multidimensional scaling (MDS) technique [*Borg and Groenen*, 1997; *Scheidt and Caers*, 2009] to visualize the variability in the ensemble of sampled models. Given a dissimilarity matrix $\mathbf{D}$ between the models, such a representation displays an ensemble of models $\mathbf{m}_i$ as a set of points in a possibly high-dimensional Euclidean space, arranged in such a way that their respective distances are preserved. $\mathbf{D}$ can be computed using any appropriate measure of distance. The coordinates of the points are in high dimension, but for representation they are projected on spaces of lower dimensionality (2D or 3D), where the distances are then only approximately preserved. In the present case, the distance between any two models $d\{\mathbf{m}_i, \mathbf{m}_j\}$ is the piecewise Euclidean distance between the heads calculated on the entire domain using both models. $\mathbf{D}$ is computed using 601 models (150 models for each of the four sampling methods, plus the reference, represented by a red dot), and each ensemble of models is represented on a different column for more clarity. In this case, representation of the points as 2D projections is adequate since the first two dimensions carry 76% of the information.

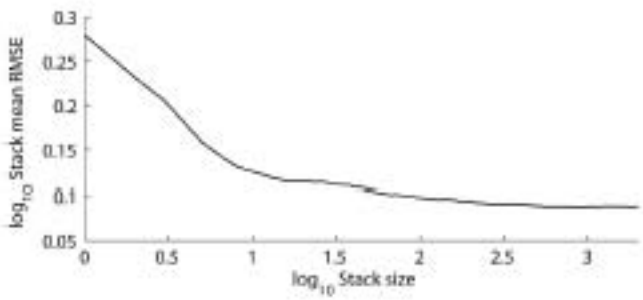


**Figure 11.** Convergence of the Metropolis sampler in the synthetic test case.

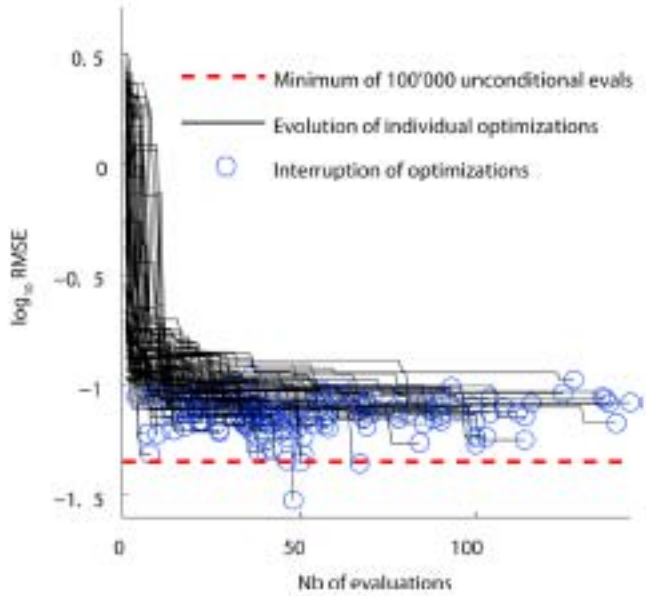


**Figure 12.** Evolution of the 150 individual optimizations used for interrupted Markov chains.

[81] The first column of Figure 9 represents the evaluation of 100,000 unconditional realizations. This ensemble characterizes the prior $f(\mathbf{m})$. On average, a large drawdown is observed at the pumping well, indicating that most of the prior models have no channel at this location. The standard deviation is large except near the boundary conditions (note that, to keep the figure readable, the standard deviation is not represented above 0.1 and heads are not represented below −1 m). Since no conditioning points data are imposed, the probability of channels is uniform. The models are very scattered in the distance space, which confirms the high variability of model responses.

[82] We start by solving this inverse problem using rejection sampling. The likelihood function used is

$$L(\mathbf{m}_i) = \exp\left(-\frac{\mathrm{RMSE}(\mathbf{m}_i)^2}{2\sigma^2}\right), \tag{6}$$

with $\sigma = 0.03$ m, which can reasonably correspond to the head measurement error. The supremum value is set to 0.607, which corresponds to a RMSE of 0.0300 (a higher fit than any of the samples). After 100,000 evaluations, 150 realizations are sampled, representative of $f(\mathbf{m}|d)$. Six of these realizations are displayed in Figure 10 with their respective fits to data. Although good fits are found, realizations are very different. This is an indication of the nonuniqueness of the solutions and of multiple local minima in the solution space.

[83] Compared to the prior models, the standard deviation of heads displays reduced uncertainty, especially at the data locations. The probability of occurrence of sand shows that the head measurements captured some essential features governing flow behavior. One such feature is the presence of a channel at the well location, slightly tilted downward and that does not branch in the immediate vicinity of the well. Another feature is the absence of channels at the location of the four observation wells close to the center. In the distance space, the posterior models represent a narrow subset of the prior. Note that more informative data, such as transient heads or concentration data, would allow a more detailed characterization of the channel locations.

[84] Now that the posterior distribution is entirely characterized with rejection, we perform another sampling using ISR as a Metropolis sampler as described in section 2.3, with likelihood (6). A constant resampling factor of $\varphi = 0.01$ is used. The chain is carried on until 2000 models are accepted. Because of the high rejection rate, 26,753 proposal solutions are evaluated in total. The convergence of the chain is displayed in Figure 11. Upon convergence,

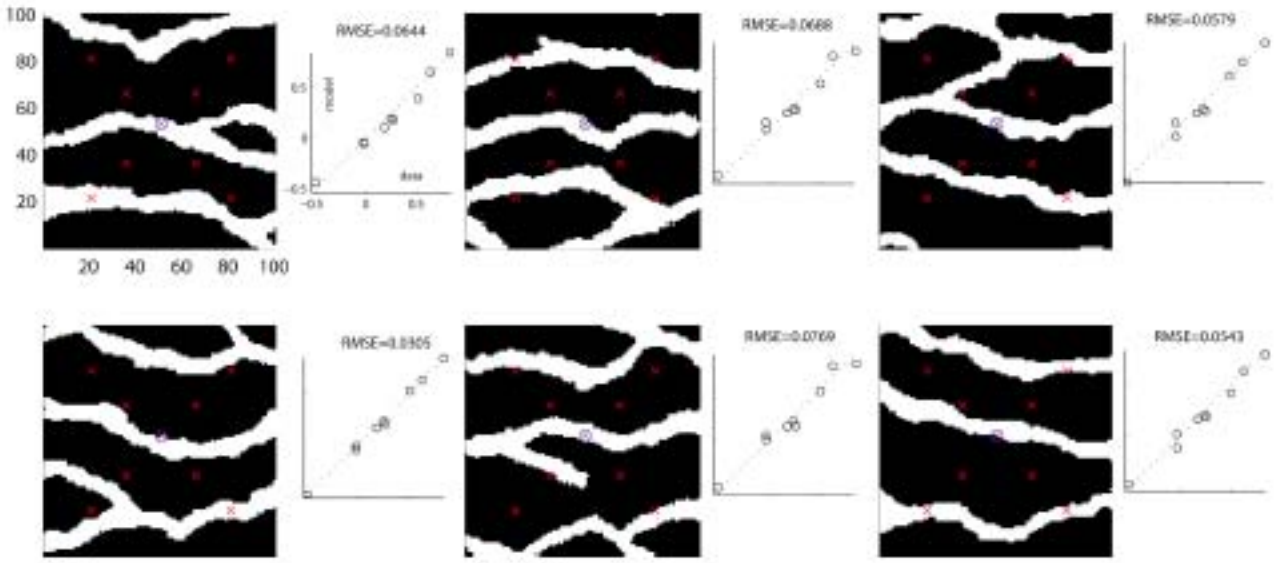


**Figure 13.** Six realizations (randomly chosen) out of the 150 sampled with interrupted Markov chains. The fits to data and the RMSE are shown on the right of each realization. The axes and labels are the same for all realizations, but they are only displayed for the top left image.

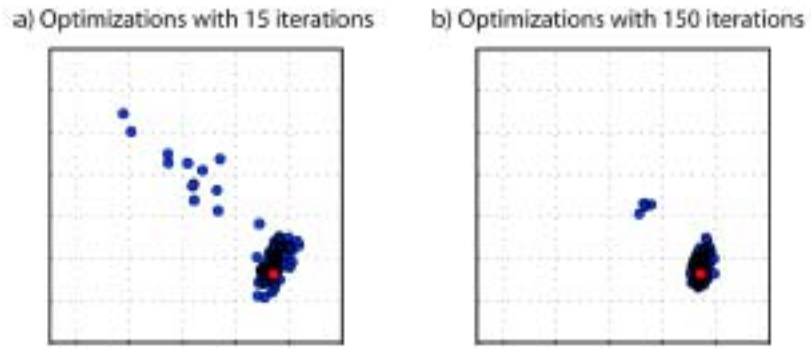


**Figure 14.** Distance-based representation of ensembles obtained with deterministic stopping criteria. (a) Fixed number of 15 iterations. (b) Fixed number of 150 iterations.

variations between the RMSEs of samples are expected to cancel out and the mean RMSE should stabilize. Using Figure 11, we define the burn-in period as the initial 200 accepted models, and we remove those from the chain. Then, one every 12 models accepted in the Markov chain is retained as a sample of the posterior distribution. As a result, 150 samples are obtained.

[85] The third column of Figure 9 shows that these 150 samples are similar to the outcomes of rejection sampling. The mean heads and the probability of occurrence of channels are fairly close to the ones obtained by rejection. Slight differences are observed for the standard deviation of heads. In the distance-based representation, both rejection and Metropolis samplers produce models that are represented in the distance space as a main cluster with a few outliers. While the main cluster is similar for both samplers, rejection produced seven outliers and Metropolis produced only four. Moreover, Metropolis sampling results in a higher median RMSE. Although both samplings are fairly similar, the differences can be attributed to the relatively small number of samples (150), but also to the imperfect conditioning of the DS simulation method.

[86] The fourth column of Figure 9 represents 150 samples obtained by interrupted Markov chains, with a constant fraction of resampled nodes of $\varphi = 0.01$. Likelihood (6) is used, and the supremum is the same as for rejection sampling. The results are relatively similar to the ensemble obtained by rejection sampling, using only 8108 forward problem evaluations (about 54 forward simulation runs for each matched model). The head standard deviation is noticeably reduced in the upper part of the image. In the distance-based representation, five models lie out of the main cluster, which is similar to what was observed with rejection sampling. However, the main cluster is too narrow (see zoomed-in part). Figure 12 shows the evolution of the 150 optimizations and their interruptions. The number of iterations $i$ before interruption ranges between 4 and 144, with an average of 54 iterations. Six optimized realizations obtained by interrupted Markov chains are shown in Figure 13. Similarly to the case of rejection sampling, the presence of diversity in the population of solutions indicates that different local minima have been explored.

[87] For comparison, Figure 14 displays the distance-based representation of models obtained with deterministic stopping criteria, after a fixed number of iterations $i_{max} = 150$ and $i_{max} = 15$. Clearly, 15 iterations are not enough and produce an ensemble that is too spread, while 150 iterations are too much, only representing a narrow subset of the desired posterior. Note that the correct number of iterations cannot be known a priori.

## 4. Conclusion

[88] We presented the iterative spatial resampling method (ISR) to perturb realizations of a spatially dependent variable while preserving its spatial structure. The method is used as a transition kernel to produce Markov chains of geostatistical realizations. Depending on the acceptance/rejection criterion in the Markov process, it is possible to obtain a chain of realizations aimed either at characterizing a certain posterior distribution with Metropolis sampling or at calibrating one realization at a time. ISR can therefore be applied in the context of Bayesian inversion or as an optimization method. For the latter case, we present a stopping criterion for optimizations inspired from importance sampling. In the studied cases, it yields posterior distributions reasonably close to the ones obtained by rejection sampling, with important reduction in CPU cost.

[89] The method is based solely on conditioning data; hence it can be used with any geostatistical technique able to produce conditional simulations. Moreover, ISR can be straightforwardly implemented without modification of existing computer codes. The method is simple in its concept and needs very little parameterization.

[90] The fraction of resampled nodes $\varphi$ is the only parameter required for optimization with ISR. It has been shown that the method is efficient for a wide range of $\varphi$. This low sensitivity is a major advantage from a practical point of view because it saves the user the hassle of performing lengthy sensitivity analysis to find optimal parameters.

[91] The approach is illustrated with both continuous and discrete variables. We use head data and groundwater flow problems, but the principle is general and can be applied to other inversion problems such as the ones involving geophysical applications. Future research will focus on extending the concept of ISR. For example, local perturbations can be obtained by resampling certain areas more than others or by using quasirandom resampling [e.g., *Tang et al.*, 2008]. This could be used when the forward problem provides local fitness or sensitivity information. Another aspect is the integration of preferential search directions. In this paper, we investigated search patterns that use random search directions, obtained by sampled locations that are not correlated between an iteration and the next one. It may be

possible to continue the search in the same direction as the previous iteration by adopting sampling locations that are dependent on the sampling at the previous iteration.

[92] Acknowledgments. This work was funded by a postdoctoral fellowship of the Swiss National Science Foundation (grant PBNEP2-124334) awarded to G.M. We thank Klaus Mosegaard for precious insights on his work and André Journel for constructive comments, as well as Andres Alcolea and two anonymous reviewers for their positive feedback.

## References

Alcolea, A., and P. Renard (2010), Blocking Moving Window sampler: Conditioning multiple-point simulations to hydrogeological data, *Water Resour. Res.*, *46*, W08511, doi:10.1029/2009WR007943.

Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, *29*(11), 1678–1689.

Alcolea, A., P. Renard, G. Mariethoz, and F. Bretone (2009), Reducing the impact of a desalination plant using stochastic modeling and optimization techniques, *J. Hydrol.*, *365*(3–4), 275–288, doi:10.1016/j.jhydrol.2008.11.034.

Ballin, P., K. Aziz, A. Journel, and L. Zaccolo (1993), Quantifying the Impact of Geological Uncertainty on Reservoir Performing Forecasts, paper presented at SPE Symposium on Reservoir Simulation, 28 February–3 March 1993, SPE, New Orleans, LA.

Bayer, P., and M. Finkel (2004), Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones, *Water Resour. Res.*, *40*, W06506, doi:10.1029/2003WR002675.

Besag, J., and C. Kooperberg (1995), On conditional and intrinsic autoregressions, *Biometrika*, *82*(4), 733–746.

Borg, I., and P. Groenen (1997), *Modern multidimensional scaling: theory and applications*, 614 pp., Springer, New York.

Caers, J. (2003), History matching under a training image-based geological model constraint, *SPE J.*, *8*(3), 218–226, SPE 74716.

Caers, J. (2005), *Petroleum Geostatistics*, 88 pp., Soc. Pet. Eng., Richardson, TX.

Caers, J. (2007), Comparing the gradual deformation with the probability perturbation method for solving inverse problems, *Math. Geol.*, *39*(1), 27–52.

Caers, J., and T. Hoffman (2006), The probability perturbation method: A new look at Bayesian inverse modeling, *Math. Geol.*, *38*(1), 81–100.

Carrera, J., and S. Neuman (1986), Estimation of aquifer parameters under transient and steady-state conditions: 2. Uniqueness, stability and solution algorithms, *Water Resour. Res.*, *22*(2), 211–227.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and J. Luit (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*, 206–222.

De Marsily, G., G. Lavedan, M. Boucher, and G. Fasanino (1984), Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Proc Geostatistics for Natural Resources Characterization. Part 2*, edited by G. Verly, pp. 831–849, D. Reidel Publishing, Dordrecht.

De Marsily, G., F. Delay, J. Gonçalvès, P. Renard, V. Teles, and S. Violette (2005), Dealing with spatial heterogeneity, *Hydrogeol. J.*, *13*(1), 161–183.

Deutsch, C., and A. Journel (1992), *GSLIB: Geostatistical Software Library*, 340 pp., Oxford Univ. Press, New York.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Emery, X. (2004), Testing the correctness of the sequential algorithm for simulating Gaussian random fields, *Stochastic Environ. Res. Risk Assess.*, *18*(6), 401–413.

Fraser, A. (1957), Simulation of genetic systems by automatic digital computers. I. Introduction, *Aust. J. Biol. Sci.*, *10*, 484–491.

Fu, J., and J. Gomez-Hernandez (2008), Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method, *Inverse Probl. Sci. Eng.*, *16*(7), 865–884.

Fu, J., and J. Gomez-Hernandez (2009), 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.

Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, 412 pp., Addison-Wesley, Berlin.

Gomez-Hernandez, J., A. Sahuquillo, and J. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. Theory, *J. Hydrol. 203*(1–4), 162–174.

Guardiano, F., and M. Srivastava (1993), Multivariate geostatistics: Beyond bivariate moments, in *Geostatistics-Troia*, edited by A. Soares, pp. 133–144, Kluwer Acad., Dordrecht, The Netherlands.

Hendricks-Franssen, H.-J., F. Stauffer, and W. Kinzelbach (2004), Joint estimation of transmissivities and recharges—application: Stochastic characterization of well capture zones, *J. Hydrol.*, *294*(1–3), 87–102.

Hendricks-Franssen, H.-J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, *Adv. Water Resour.*, *32*(6), 851–872, doi:10.1016/j.advwatres.2009.02.011.

Hernandez, A., S. Neuman, A. Guadagnini, and J. Carrera (2006), Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, *42*, W05425, doi:10.1029/2005WR004449.

Hu, L. (2000), Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Math. Geol.*, *32*(1), 87–108.

Hu, L., and T. Chugunova (2008), Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resour. Res.*, *44*, W11413, doi:10.1029/2008WR006993.

Hu, L., and M. Le Ravalec-Dupin (2004), On some controversial issues of geostatistical simulation, *Geostatistics Banff 2004*, Kluwer Acad., Banff, AB.

Hu, L., G. Blanc, and B. Noetinger (2001), Gradual deformation and iterative calibration of sequential stochastic simulations, *Math. Geol.*, *33*(4), 475–489.

Johansen, K., J. Caers, and S. Suzuki (2007), Hybridization of the probability perturbation method with gradient information, *Comput. Geosci.*, *11*(4), 319–331.

Journel, A., and T. Zhang (2006), The necessity of a multiple-point prior model, *Math. Geol.*, *38*(5), 591–610.

Karpouzos, D., F. Delay, K. Katsifarakis, and G. De Marsily (2001), A multipopulation genetic algorithm to solve the inverse problem in hydrogeology, *Water Resour. Res.*, *37*(9), 2291–2302.

Kirkpatrick, S., C. Gelatt, and M. Vecchi (1983), Optimization by simulated annealing, *Science*, *220*(4598), 671–680.

Kitanidis, P. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411–2419.

Kjønsberg, H., and O. Kolbjørnsen (2008), Markov mesh simulations with data conditioning through indicator kriging, in *Proceedings of the Eighth International Geostatistics Congress 2008*, 1–5 Dec. 2008, edited by J. M. Ortiz and X. Emery, pp. 257–267, Gecamin Ltd., Santiago.

Le Ravalec-Dupin, M. (2010), Pilot block method methodology to calibrate stochastic permeability fields to dynamic data, *Math. Geosci.*, *42*(2), 165–185, doi:10.1007/s11004-009-9249-x.

Le Ravalec-Dupin, M., and L. Hu (2007), Combining the pilot point and gradual deformation methods for calibrating permeability models to dynamic data, *Oil Gas Sci. Technol.*, *62*(2), 169–180.

Le Ravalec-Dupin, M., and B. Noetinger (2002), Optimization with the gradual deformation method, *Math. Geol.*, *34*(2), 125–142.

Liu, N., and D. Oliver (2004), Experimental assessment of gradual deformation method, *Math. Geol.*, *36*(1), 65–77.

Liu, X., A. Cardiff, and K. Kitanidis (2010), Parameter estimation in nonlinear environmental problems, *Stochastic Environ. Res. Risk Assess.*, 1436–3240, doi:10.1007/s00477-010-0395-y.

Llopis-Albert, C., and L. Cabrera (2009), Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 3. Application to the Macrodispersion Experiment (MADE-2) site, on Columbus Air Force Base in Mississippi (USA), *J. Hydrol.*, *371*(1–4), 75–84.

Mariethoz, G. (2010), A general parallelization strategy for random path based geostatistical simulation methods, *Comput. Geosci.*, *37*(7), 953–958, doi:10.1016/j.cageo.2009.11.001.

Mariethoz, G., and P. Renard (2010), Reconstruction of incomplete data sets or images using Direct Sampling, *Math. Geosci.*, *42*(3), 245–268, doi:10.1007/s11004-010-9270-0.

Mariethoz, G., P. Renard, F. Cornaton, and O. Jaquet (2009), Truncated plurigaussian simulations to characterize aquifer heterogeneity, *Ground Water*, *47*(1), 13–24, doi:10.1111/j.1745-6584.2008.00489.x.

Mariethoz, G., P. Renard, and J. Straubhaar (2010), The direct sampling method to perform multiple-point simulations, *Water Resour. Res.*, doi:10.1029/2008WR007621, in press.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, and A. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12431–12447.

Oliver, D., L. Cunha, and A. Reynolds (1997), Markov chain Monte Carlo methods for conditioning a logpermeability field to pressure data, *Math. Geosci.*, *29*(1), 61–91.

Omre, H., and H. Tjelmeland (1996), *Petroleum Geostatistics*, Dept. of Math. Sci., Norwegian Univ. of Sci. and Technol., Trondheim, Norway.

Pan, L., and L. Wu (1998), A hybrid global optimization method for inverse estimation of hydraulic parameters: Annealing-simplex method, *Water Resour. Res.*, *34*(9), 2261–2269.

RamaRao, B., A. LaVenue, G. De Marsily, and M. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments, *Water Resour. Res.*, *31*(3), 475–493.

Remy, N., A. Boucher, and J. Wu (2009), *Applied Geostatistics with SGeMS: A User's Guide*, 284 pp., Cambridge Univ. Press, Cambridge, U. K.

Ronayne, M., S. Gorelick, and J. Caers (2008), Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach, *Water Resour. Res.*, *44*, W08426, doi:10.1029/2007WR006635.

Scheidt, C., and J. Caers (2009), Representing spatial uncertainty using distances and kernels, *Math. Geosci.*, *41*(4), 397–419.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 379–423.

Smith, J. (1997), Quick simulation: A review of importance sampling techniques in communications systems, *IEEE J. Sel. Areas Commun.*, *15*(4), 597–613.

Straubhaar, J., P. Renard, G. Mariethoz, R. Froidevaux, and O. Besson (2010), An improved parallel multiple-point algorithm using a list approach, *Math. Geosci.*, in press.

Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.*, *34*(1), 1–22.

Subbey, S., M. Christie, and M. Sambridge (2004), Prediction under uncertainty in reservoir modeling, *J. Pet. Sci. Eng.*, *44*(1–2), 143–153.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2008), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817, doi:10.5194/hess-11-793-2007.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Parameter estimation*, Soc. Ind. Appl. Math., Philadelphia, Pa.

Tjelmeland, H., and J. Besag (1998), Markov random fields with higher-order interactions, *Scand. J. Stat.*, *25*, 415–433.

Vesselinov, V., S. Neuman, and W. Illman (2001), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff. 1. Methodology and borehole effects, *Water Resour. Res.*, *37*(12), 3001–3017.

von Neumann, J. (1951), Various techniques used in connection with random digits. Monte Carlo methods, *Natl. Bur. Stand.*, *12*, 36–38.

Yeh, W. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(2), 95–108.

Zanini, A., and K. Kitanidis (2009), Geostatistical inversing for large-contrast transmissivity fields, *Stochastic Environ. Res. Risk Assess.*, *23*(5), 656–577.

Zheng, C., and P. Wang (1996), Parameter structure identification using tabu search and simulated annealing, *Adv. Water Resour.*, *19*(4), 215–224.

Zimmerman, D., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*(6), 1373–1413.

---

J. Caers, ERE Department, Stanford University, 367 Panama St., Rm. 65, Stanford, CA 94305-2220, USA.

G. Mariethoz, National Centre for Groundwater Research and Training, The University of New South Wales, Anzac Parade, Sydney, NSW 2033, Australia. (gregoire.mariethoz@unsw.edu.au)

P. Renard, Centre for Hydrogeology, University of Neuchâtel, 11 Rue Emile Argand, CP 158, CH-2009 Neuchâtel, Switzerland.