

# 1 Bayesian inverse problem and optimization with iterative 2 spatial resampling

3 Grégoire Mariethoz,<sup>1,2</sup> Philippe Renard,<sup>1</sup> and Jef Caers<sup>2</sup>

4 Received 8 March 2010; revised 4 July 2010; accepted 17 August 2010; published XX Month 2010.

5 [1] Measurements are often unable to uniquely characterize the subsurface at a desired  
6 modeling resolution. In particular, inverse problems involving the characterization of  
7 hydraulic properties are typically ill posed since they generally present more unknowns than  
8 data. In a Bayesian context, solutions to such problems consist of a posterior ensemble of  
9 models that fit the data (up to a certain precision specified by a likelihood function) and that  
10 are a subset of a prior distribution. Two possible approaches for this problem are Markov  
11 chain Monte Carlo (McMC) techniques and optimization (calibration) methods. Both  
12 frameworks rely on a perturbation mechanism to steer the search for solutions. When the  
13 model parameters are spatially dependent variable fields obtained using geostatistical  
14 realizations, such as hydraulic conductivity or porosity, it is not trivial to incur perturbations  
15 that respect the prior spatial model. To overcome this problem, we propose a general  
16 transition kernel (iterative spatial resampling, ISR) that preserves any spatial model  
17 produced by conditional simulation. We also present a stochastic stopping criterion for  
18 the optimizations inspired from importance sampling. In the studied cases, this yields  
19 posterior distributions reasonably close to the ones obtained by a rejection sampler, but  
20 with a greatly reduced number of forward model runs. The technique is general in the sense  
21 that it can be used with any conditional geostatistical simulation method, whether it  
22 generates continuous or discrete variables. Therefore it allows sampling of different priors  
23 and conditioning to a variety of data types. Several examples are provided based on either  
24 multi-Gaussian or multiple-point statistics.

25 **Citation:** Mariethoz, G., P. Renard, and J. Caers (2010), Bayesian inverse problem and optimization with iterative spatial  
26 resampling, *Water Resour. Res.*, 46, XXXXXX, doi:10.1029/2010WR009274.

## 27 1. Introduction

28 [2] Integrating state variables in hydrogeological site  
29 characterization by solving an inverse problem continues  
30 to be an important topic of investigation [Carrera *et al.*,  
31 2005; Hendricks-Franssen *et al.*, 2009; Liu *et al.*, 2010;  
32 Zimmerman *et al.*, 1998]. Indeed, inverse problems are a  
33 crucial aspect of groundwater modeling since they are  
34 used to validate or invalidate certain geological scenarios  
35 [Ronayne *et al.*, 2008] as well as to reduce model uncer-  
36 tainty for engineering prediction and decision making prob-  
37 lems [Alcolea *et al.*, 2009]. As such, models need to not  
38 just match the data, they also need to be predictive, a  
39 property that is difficult to objectively verify [Ballin *et al.*,  
40 1993; Subbey *et al.*, 2004]. Conditioning models to points  
41 data (localized measurements of the variable of interest) is  
42 addressed very efficiently by most geostatistical simulation  
43 algorithms [Deutsch and Journel, 1992; Remy *et al.*, 2009].  
44 In this paper, we refer to conditioning models to indirect  
45 state variable data (such as heads).

46 [3] Problems involving flow in underground media typi-  
47 cally present more unknowns than data. For example, 48  
48 modeling hydraulic conductivity or porosity on an entire 49  
49 domain, based only on local head measurements or tracer 50  
50 tests, is typically an ill-posed inverse problem [Carrera 51  
51 *et al.*, 2005; De Marsily *et al.*, 2005; Yeh, 1986]. Ill- 52  
52 posedness means that multiple solutions are possible, and 53  
53 characterizing the uncertainty spanned by these multiple so- 54  
54 lutions is often critical in real field engineering use of these 55  
55 models. Other consequences of ill-posedness can be that a 56  
56 solution does not exist or is unstable with regard to small 57  
57 variations in the input data [Carrera and Neuman, 1986]. In a 58  
58 Bayesian framework, these issues are dealt with by obtaining 59  
59 a posterior distribution given a certain prior distribution and a 60  
60 likelihood function. In this respect, only Markov chain Monte 61  
61 Carlo (McMC) methods have been shown to sample with 62  
62 reasonable accuracy from this posterior [Mosegaard and 63  
63 Tarantola, 1995; Omre and Tjelmeland, 1996], i.e., to gen- 64  
64 erate model realizations that (1) match the points data and the 65  
65 indirect state data, (2) reproduce for each inverse solution 66  
66 some prior statistics (e.g., a spatial covariance) and (3) sample 67  
67 correctly from the posterior as imposed by Bayes' rule. Most 68  
68 gradient-based/optimization techniques [De Marsily *et al.*, 69  
69 1984; Gomez-Hernandez *et al.*, 1997; Hernandez *et al.*, 69  
70 2006; RamaRao *et al.*, 1995; Vesselinov *et al.*, 2001] do not 71  
71 completely fulfill these three requirements.

<sup>1</sup>Centre for Hydrogeology, University of Neuchâtel, Neuchâtel, Switzerland.

<sup>2</sup>ERE Department, Stanford University, Stanford, California, USA.

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

[11] 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  $Q(\mathbf{m}_i|\mathbf{m}^*) = Q(\mathbf{m}^*|\mathbf{m}_i)$ .

[12] 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

186 Monte Carlo (BMcMC) method that incurs local pertur-  
 187 bations by successively resimulating a square area of the  
 188 realizations (a block). The BMcMC method has been used  
 189 for sampling the posterior distribution of synthetic inverse  
 190 problems in a multi-Gaussian framework [*Fu and Gomez-*  
*Hernandez, 2009*].

192 [13] Optimization methods aim at finding realizations that  
 193 maximize the likelihood  $f(\mathbf{d}|\mathbf{m})$ . They do not allow char-  
 194 acterizing  $f(\mathbf{m}|\mathbf{d})$  but are often used since they are much  
 195 more efficient than sampling algorithms. These methods  
 196 repeatedly update an initial solution to minimize an objec-  
 197 tive function, often measuring a misfit to measured data.  
 198 Although regularization terms can be added to make the  
 199 perturbed models look more realistic, prior constraints are  
 200 often minimal. One can use either gradient-based [*Carrera*  
*and Neuman, 1986*] or gradient-free [*Karpouzos et al.,*  
*2001 2001*] methods. Since they search in a stochastic manner,  
 203 gradient-free methods are less prone to be trapped in local  
 204 minima (i.e., it is guaranteed that the global minimum is  
 205 found after an infinite number of iterations). Upon conver-  
 206 gence, a single calibrated solution is obtained. When several  
 207 local minima are present, one can obtain alternative solu-  
 208 tions by repeating the optimization procedure using different  
 209 starting points.

210 [14] Among the optimization techniques, simulated an-  
 211 nealing [*Kirkpatrick et al., 1983*] has been extensively used  
 212 to solve inverse groundwater modeling problems [e.g., *Pan*  
*213 and Wu, 1998; Zheng and Wang, 1996*]. Genetic algo-  
 214 rithms [*Fraser, 1957; Goldberg, 1989*] have been used for  
 215 identifying structures in hydraulic conductivity fields  
 216 [*Karpouzos et al., 2001*]. *Alcolea and Renard [2010]* apply  
 217 the BMcMC method and simulated annealing to optimize  
 218 non-multi-Gaussian random fields generated using the  
 219 *impala* multiple-point simulation code (J. Straubhaar et al.  
 220 An improved parallel multiple-point algorithm, submitted to  
 221 *Mathematical Geosciences*, 2010).

222 [15] The gradual deformation method (GDM) [*Hu, 2000;*  
*223 Hu et al., 2001*] and the probability perturbation method  
 224 (PPM) [*Caers, 2003; Caers and Hoffman, 2006; Johansen*  
*225 et al., 2007*] proceed by combining uniformly sampled  
 226 realizations. By adjusting a single parameter, they allow  
 227 obtaining a smooth transition from one simulation to another  
 228 while preserving a prior structural model. Therefore, finding  
 229 a calibrated realization can be accomplished by a series of  
 230 1D optimizations. These methods have been successfully  
 231 applied in hydrogeology and petroleum engineering [*Le*  
*232 Ravalec-Dupin and Hu, 2007; Le Ravalec-Dupin, 2010;*  
*233 Llopis-Albert and Cabrera, 2009; Ronayne et al., 2008*].

234 [16] In this paper we present a transition kernel (iterative  
 235 spatial resampling, ISR) that can be used either as a proposal  
 236 distribution with McMC sampling methods or as a perturba-  
 237 tion strategy when optimization is used. We show that, in both  
 238 cases, it yields accurate sampling of the posterior. To validate  
 239 the results, we use the rejection sampler as a reference.

## 240 2.2. Reproduction of the Prior

241 [17] A premise of both rejection and Metropolis samplers  
 242 is that proposal models  $\mathbf{m}^*$  have a nonzero prior probability.  
 243 It is not possible for a rejection sampler to produce samples  
 244 with zero prior probability. We will define as a bounded  
 245 prior a prior probability density for which there exist models  
 246  $m$  that have zero probability density. For example, in a

Boolean model where simulated objects all exhibit the same 247  
 (deterministic) direction, any model with different object 248  
 directions has zero probability of occurrence. On the other 249  
 hand, examples of unbounded prior densities are the multi- 250  
 Gaussian model and the Markov random field model [*Besag* 251  
*and Kopperberg, 1995; Tjelmeland and Besag, 1998*]. With 252  
 unbounded priors, all models are possible, even those with 253  
 extremely low prior density. For example, in the standard 254  
 multi-Gaussian, the model  $m = 0$  (a model with zeros 255  
 everywhere and hence variance = 0) is highly improbable 256  
 but not impossible. Actually, this very small probability can 257  
 be regarded as a mathematical artifact because the chance of 258  
 sampling it is negligible. 259

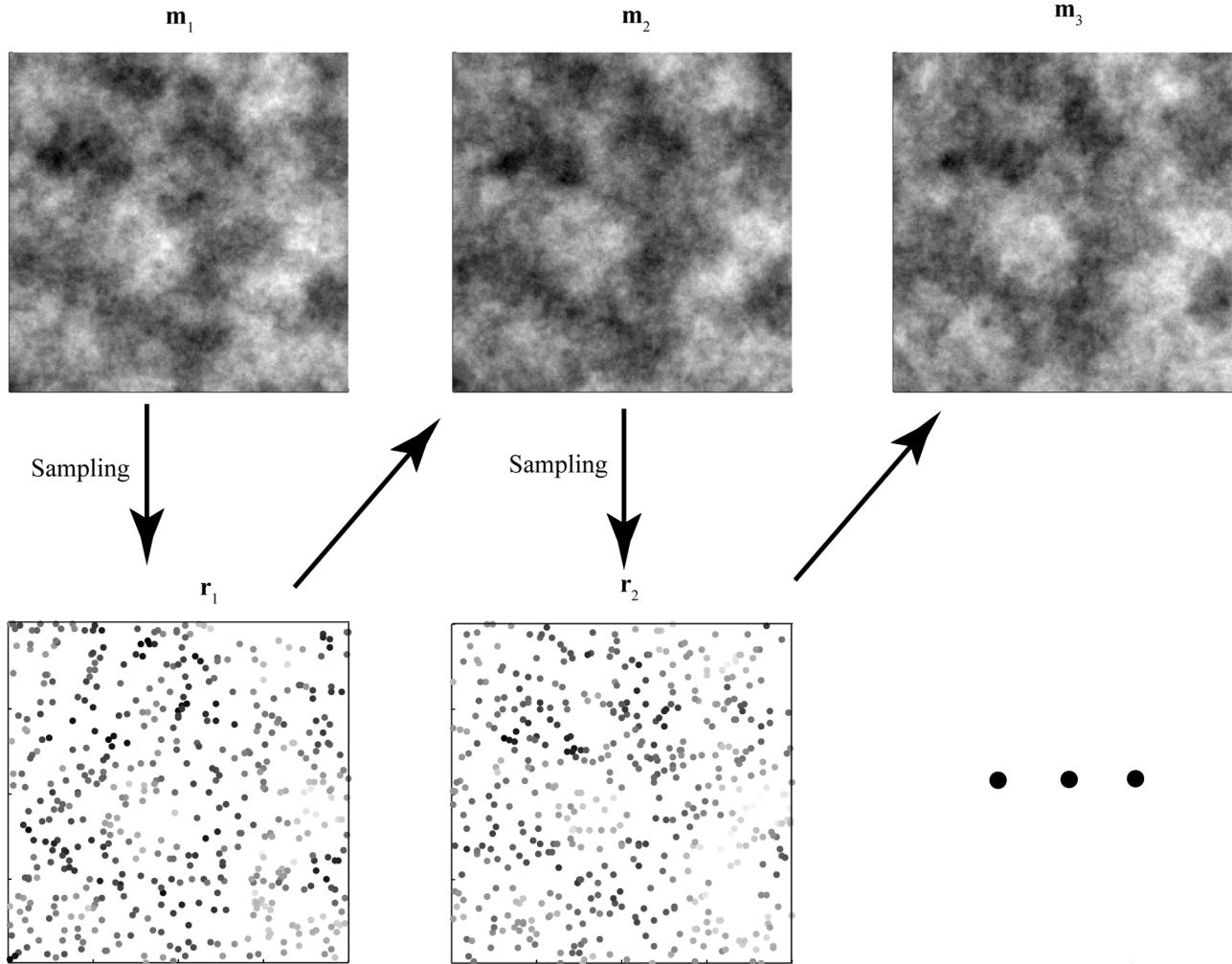
[18] Application of Bayes' rule (equation (1)) assumes 260  
 that an intersection between prior and likelihood exists. In 261  
 other terms, if all samples that have a (practically) nonzero 262  
 prior probability also have a (practically) zero likelihood, 263  
 this means that prior and likelihood are incompatible and the 264  
 solution is nonidentifiable [*Carrera and Neuman, 1986*]. 265  
 The posterior is then undetermined since it results, after 266  
 normalization, in a division of zero by zero. In such cases, 267  
 prior and likelihood do not intersect, Bayes' rule is not 268  
 applicable, and a modeling decision has to be taken whether 269  
 to put in question the data, the likelihood model, or the 270  
 prior. This decision should be motivated by reconsidering 271  
 the basis adopted to define the prior as well as the confi- 272  
 dence given to the data. 273

[19] In the absence of intersection between prior and like- 274  
 lihood, data-driven inverse modeling techniques match the 275  
 data at the price of producing models with zero or very low 276  
 prior probability (for example, not preserving the specified 277  
 spatial dependence). On the other hand, prior-driven techni- 278  
 ques favor the prior and hence the statistics reflected in prior 279  
 models (such as a variogram), but may be unable to achieve 280  
 good fits to conflicting data. By opting for a certain inverse 281  
 modeling technique, the modeler decides which piece of 282  
 information should prevail in case of incompatibility. 283

[20] Among data-driven methods, we mention the gradual 284  
 deformation method (GDM) [*Caers, 2007; Hu and Le*  
*285 Ravalec-Dupin, 2004; Le Ravalec-Dupin and Noetinger,*  
*286 2002; Liu and Oliver, 2004*], the quasi-linear method 287  
 [*Kitanidis, 1995; Zanini and Kitanidis, 2009*], and the reg- 288  
 ularized pilot points method (RPPM) that considers con- 289  
 straints imposed by both data and prior and adjusts these 290  
 constraints using a weighted regularization term [*Alcolea* 291  
*et al., 2006; Doherty, 2003; Hendricks-Franssen et al.,*  
*292 2004*]. In the class of prior-driven methods, one can find 293  
 the probability perturbation method [*Caers and Hoffman,* 294  
*2006*] and the blocking moving window [*Alcolea and*  
*295 Renard, 2010*]. ISR, the technique we present in this paper, 296  
 also belongs to this class of prior-driven methods. 297

## 298 2.3. Iterative Spatial Resampling

[21] Let a model  $\mathbf{m}_i = \{Z_i(\mathbf{x}_1), \dots, Z_i(\mathbf{x}_M)\}$  be a realization 299  
 of a random variable  $Z$  discretized on a grid with  $M$  nodes. 300  
 Unconditional realizations of  $Z$  are considered samples of 301  
 the prior, whether this prior is explicitly stated such as is the 302  
 case of a multi-Gaussian model or whether this prior is 303  
 defined by a given stochastic algorithm with a set of para- 304  
 meters (A. Boucher, Algorithm-driven and representa- 305  
 driven random function: A new formalism for applied 306



**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.

307 geostatistics, edited, Stanford Center for Reservoir Fore-  
308 casting, Palo Alto, CA, 2007).

309 [22] To implement sampling and searching strategies, one  
310 needs to create a chain of dependant realizations. Conse-  
311 quently, one wants to draw proposal models  $\mathbf{m}^*$  not from  
312  $f(\mathbf{m})$  but from  $Q(\mathbf{m}|\mathbf{m}_i)$ , where  $\mathbf{m}_i$  is the previous model  
313 in the chain. To preserve the spatial continuity defined by  
314 the geostatistical simulation algorithm, the conditional term  
315 should ideally be incorporated in the method used to gen-  
316 erate the realizations. Since most simulation methods also  
317 allow generating realizations conditioned to points data, we  
318 propose to use this conditioning capability to impose a  
319 conditional term on the prior. More specifically, dependence  
320 between  $\mathbf{m}^*$  and  $\mathbf{m}_i$  is introduced by extracting a subset of  
321 realization  $\mathbf{m}_i$  as  $n$  randomly located points  $\mathbf{r}_i = \{Z_i(\mathbf{x}_\alpha),$   
322  $\alpha = 1, \dots, n\}$  and to impose these points as conditioning  
323 data to generate  $\mathbf{m}^*$ . The amount  $n$  is a tuning parameter.  
324 Proposal models are drawn from  $f(\mathbf{m})$ , but at the same  
325 time they depend on  $\mathbf{r}_i$ , itself a subset of  $\mathbf{m}_i$ .  
326 [23] Creating a Markov chain using ISR is accomplished  
327 by performing the following steps:

- [24] 1. Generate an initial model  $\mathbf{m}_1$  using a geostatistical simulation algorithm, and evaluate its likelihood  $L(\mathbf{m}_1)$ . 328
- [25] 2. Iterate on  $i$ : 329
- [26] a. Select randomly a subset  $\mathbf{r}_i = \{Z_i(\mathbf{x}_\alpha), \alpha = 1, \dots, n\}$  330
- of  $n$  points belonging to  $\mathbf{m}_i$ . 331
- [27] b. Generate a proposal realization  $\mathbf{m}^*$  by conditional 332 simulation using  $\mathbf{r}_i$  and the same geostatistical model with a 333 new random seed. 334
- [28] c. Evaluate  $L(\mathbf{m}^*)$ . 335
- [29] d. Accept or reject  $\mathbf{m}^*$ . If accepted, set  $\mathbf{m}_{i+1} = \mathbf{m}^*$ , 336 otherwise go back to a (i.e., do not increment  $i$ ). If the 337 acceptance criterion is the one proposed by *Mosegaard and* 338 *Tarantola* [1995], the chain is a Metropolis sampler. 339
- [30] The method is illustrated in Figure 1, where an initial 340 sequential Gaussian simulation (SGS) realization is iter- 341 atively perturbed. However, any simulation method can be 342 used, as long as it is able to produce conditional simulations. 343 Using ISR when actual conditioning points data are present 344 (for example, corresponding to field measurements) can be 345 accomplished in a straightforward manner by adding, at each 346 iteration, the “real” conditioning data to the sampled set  $\mathbf{r}$ . 347

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

[32] 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 363 points data.

[33] 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.

[34] 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

[35] 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.

[36] In theory, the proposal distribution  $Q(\mathbf{m}_i | \mathbf{m}^*)$  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 (Boucher, edited manuscript, 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].

[37] 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.

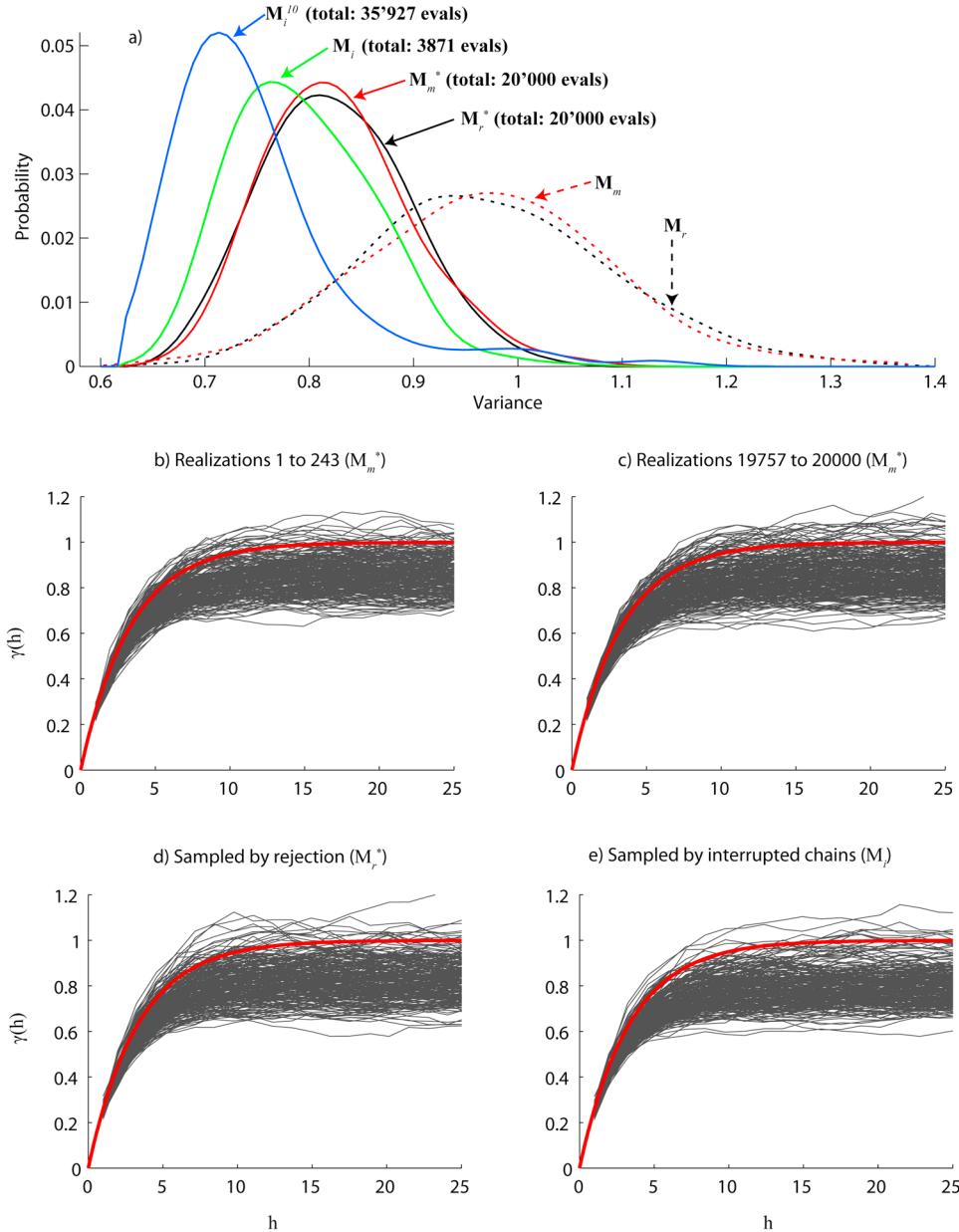
[38] 1. We characterize the prior numerically by generating an ensemble  $\mathbf{M}_r$  of 20,000 unconditional realizations, uniformly sampled from  $f(\mathbf{m})$ , and observe the variance of the simulated values (Figure 2a, black dashed line). The variance ranges approximately between 0.7 and 1.3.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

[39] 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.

[40] 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

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

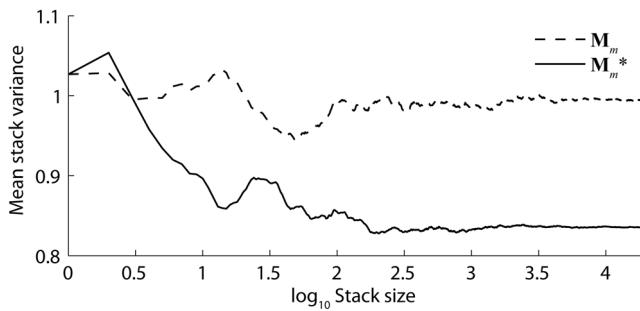
We set  $\sigma = 0.2$  so that the likelihood quickly decreases with larger variances and approaches 0 when the variance equals the prior value of 1. This constraint of minimal variance is not compatible with the prior that imposes a unit variance, although fluctuations are possible within a certain 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.

[41] 3. Similarly to experiment 1, we want to accurately  
469 characterize the posterior distribution of the problem by  
470 rejection sampling. We apply rejection sampling to the  
471 realizations of  $\mathbf{M}_r$ , with the supremum  $L(\mathbf{m})_{\max}$  chosen such  
472 that it corresponds to a variance of 0.6, and 243 samples are  
473 accepted that constitute the reference posterior ensemble  
474  $\mathbf{M}^*$ , representative of  $f(\mathbf{m}|\mathbf{l}, \mathbf{d})$  (Figure 2a, solid black line).  
475 [42] 4. For the last numerical experiment, we perform a  
476 chain with the acceptance criterion of Mosegaard and Tar-  
477 antola and the likelihood equation (3). Ideally, it should  
478 converge to  $\mathbf{M}^*$ . After a burn-in period ensuring that steady  
479 state of the Markov chain is reached, we perform iterations  
480 until an ensemble  $\mathbf{M}_m^*$  of 20,000 samples is obtained,

and we keep only one every  $m = 100$  accepted models 481  
(Figure 2a, red solid line). Figure 3 shows that the number 482  
of iterations is large enough to ensure convergence. The 483  
variance distributions of  $\mathbf{M}_r^*$  and  $\mathbf{M}_m^*$  are similar, showing 484  
that the posterior is accurately sampled. To make sure 485  
that the prior spatial model is preserved, we compare in 486  
Figures 2b and 2c the experimental variograms of the rea- 487  
lizations at the beginning and end of the Markov chain. 488  
Variogram shapes and ranges are preserved throughout the 489  
iterations, and therefore the Markov chain did not drift 490  
toward models with a (practically) zero prior probability. 491  
Note that the variograms present lower sills (smaller var- 492  
iances) than the variogram model used as input for SGS (red 493



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

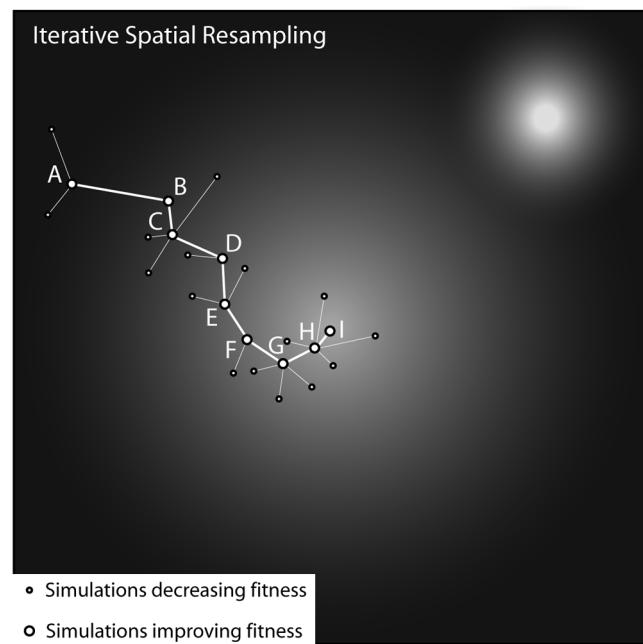
494 line). Indeed, the posterior models are not the same as the  
495 prior models since the inclusion of data yields a different  
496 ensemble. This is coherent with Bayes' rule that acknowl-  
497 edges the influence of the likelihood. In this case, prior  
498 models with a low variance have been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

### 506 2.5. Using ISR for Optimization

507 [43] Mosegaard and Tarantola [1995] indicate that their  
508 sampling method can also be used for optimization. To this  
509 end, one can create a chain of ever-improving realizations  
510 using for acceptance criterion (step 2d of the ISR algorithm  
511 as described in section 2.3):

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

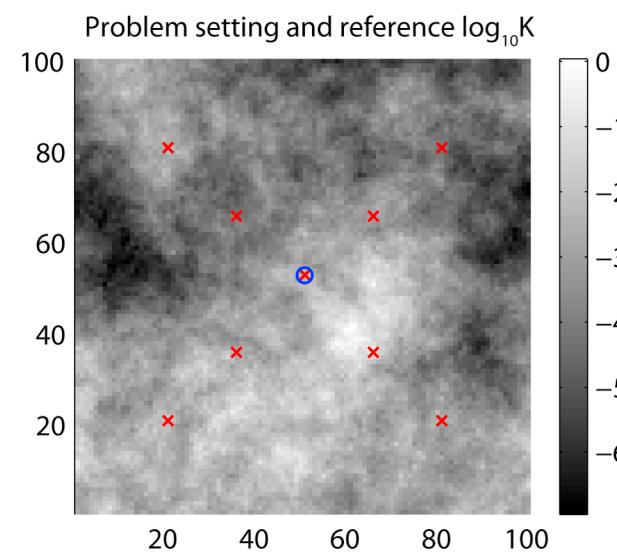
512 [44] The resulting McMC process is a stochastic search  
513 for a single calibrated model. The search strategy of ISR  
514 performs by successive steps in random directions and of  
515 random size (step size is random, but its distribution is  
516 controlled by  $\varphi$ ). When large steps occur, it allows explor-  
517 ing various regions of the solution space. Large steps are  
518 also an opportunity to leap out of local minima. On the other  
519 hand, the occurrence of small steps allows fine tuning  
520 suboptimal solutions. Since the search is stochastic, the  
521 global minimum will eventually be reached after an infinite  
522 number of iterations. However, in most practical applica-  
523 tions, it will remain in a local minimum (i.e., suboptimal).  
524 Figure 4 schematically depicts how a local minimum is  
525 searched in a simple 2D solution space. The background  
526 image represents the real, unknown solution space, with a  
527 local minimum in the center of the image and the global  
528 minimum in the top right. In this case, the search remains in  
529 the local minimum because criterion (4) is used and the  
530 number of iterations is finite. Since (4) only considers  
531 the rank of a proposal solution compared to a previous one,  
532 the search is similar to the minimization of an objective  
533 function. In this sense optimization with ISR is similar  
534 to evolutionary strategies where probabilistically generated  
535 individuals are sequentially improved while only the best  
536 one is preserved [Bayer and Finkel, 2004]. The likelihood of



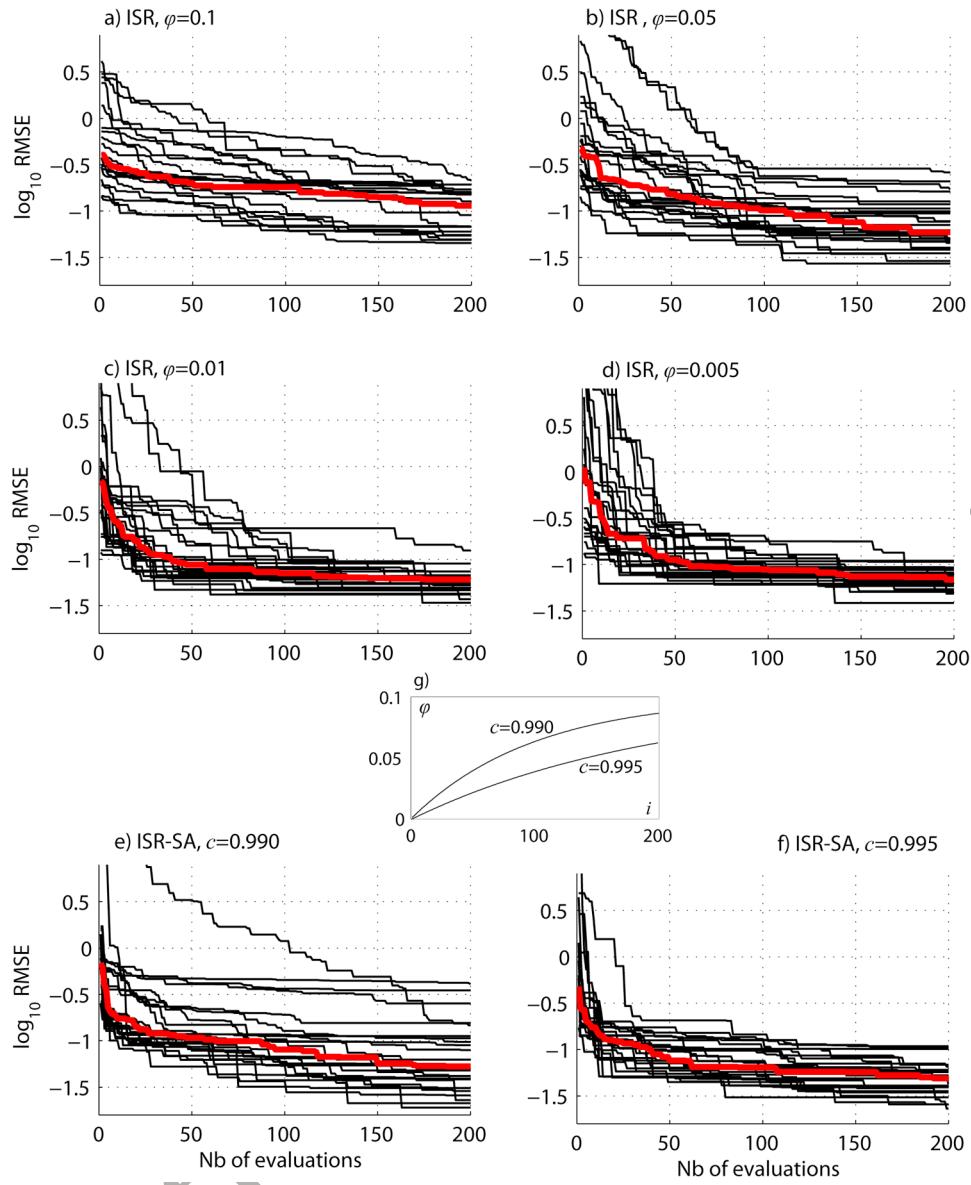
**Figure 4.** Schematic representation of the search strategy used by ISR in a simple solution space having two degrees of freedom. The background color represents the actual unknown solution space, with values ranging from black (bad solution) to white (good solution). A local minimum lies in the center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

the final solution depends on algorithmic parameters such as 537 the stopping criterion used. 538

[45] So far a constant fraction of resampled nodes has 539 been considered, but alternatives can be envisioned for 540 optimization. For example,  $\varphi$  can increase at each iteration  $i$ . 541 The optimization starts with large steps (i.e., the solution is 542 largely perturbed at the exploration phase) and finishes 543



**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  $\varphi$ . 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.

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

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

546 [46] The larger  $c$  is, the slower  $\varphi(i)$  will reach  $\varphi_{\max}$ . This  
547 is similar to simulated annealing, with the parameter  $c$   
548 defining the cooling schedule and  $\varphi_{\max}$  is the maximum  
549 value of  $\varphi$  (after an infinite number of iterations). Yet, alike  
550 other simulated annealing algorithms, adjusting  $c$  can be  
551 tedious. Nevertheless, we will see below that using (5) can  
552 accelerate the convergence compared to keeping  $\varphi$  constant.  
553 Note that a varying  $\varphi$  cannot be used in the context of a  
554 Bayesian inverse problem (Metropolis acceptance criterion)  
555 because it would make the proposal density nonsymmetric.  
556 In the context of optimization, (5) is applicable since the  
557 problem is not to sample the posterior, but to reach a local  
558 minima quickly.

559

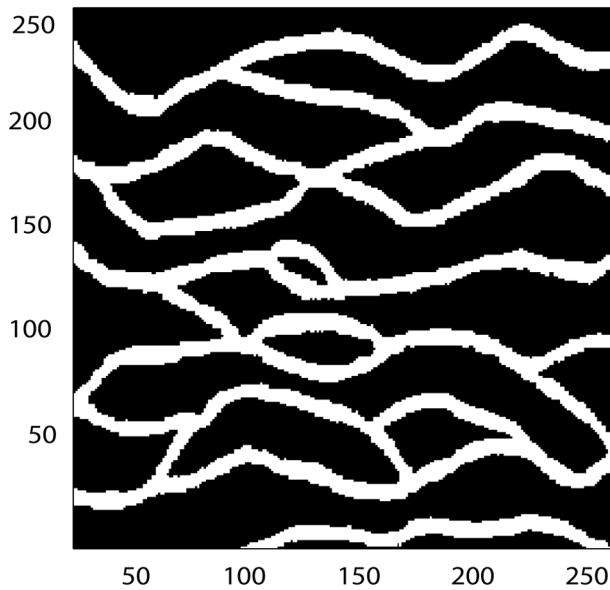
## 2.6. Sensitivity to $\varphi$

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

564

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

	minimum RMSE	median RMSE	maximum RMSE	t1.1
ISR, $\varphi = 0.1$	0.0451	0.1141	0.2124	t1.3
ISR, $\varphi = 0.05$	0.0271	0.0591	0.2610	t1.4
ISR, $\varphi = 0.01$	0.0340	0.0607	0.1242	t1.5
ISR, $\varphi = 0.005$	0.0384	0.0685	0.1096	t1.6
ISR-SA, $c = 0.990$	0.0190	0.0529	0.4094	t1.7
ISR-SA, $c = 0.995$	0.0231	0.0481	0.1087	t1.8



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

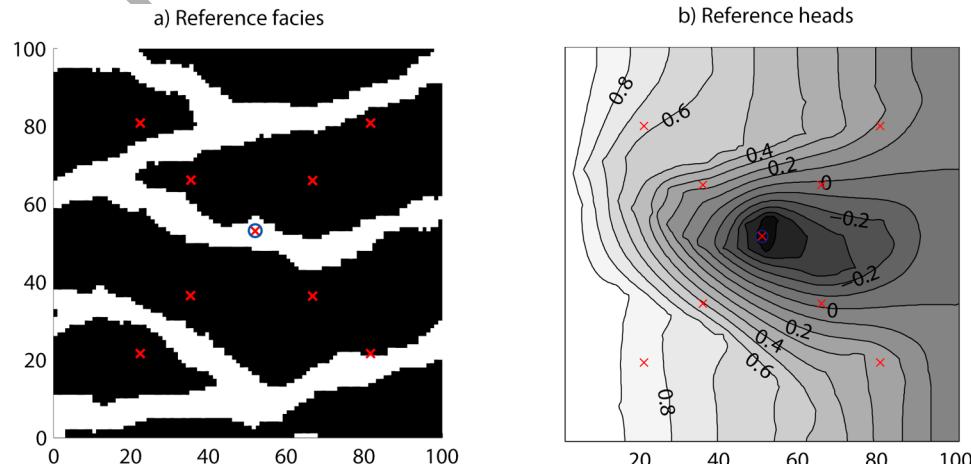
[48] The problem setting consists of a square aquifer of 566 100 m by 100 m, discretized in 100 by 100 elements 567 (Figure 5). A reference log  $K$  field is generated using SGS. 568 The spatial model for hydraulic conductivity is multi- 569 Gaussian, with an isotropic exponential variogram of range 570 90 m. The mean logarithm of the hydraulic conductivity is 571 -3, with a variance of 2. The upper and lower sides of the 572 model are no-flow boundaries, a fixed head boundary 573 condition of 1 m is set on the left side, and a fixed head 574 boundary condition of 0 m is set on the right side. A 575 pumping well extracting 0.003 m<sup>3</sup>/s is set in the center of 576 the domain (blue circle), and nine observation wells are 577 positioned at the locations of the red crosses in Figure 5 (the 578 pumping well is also an observation well). The problem is 579 solved in steady state. The nine head measurements are the 580 only constraints used to solve the inverse problem, i.e., to 581

find log  $K$  fields that match these data. We do not impose 581 any conditioning points data in order to not over-constrain 582 the prior. 583

[49] Proposal solutions are generated using SGS with the 584 true variogram model. The acceptance criterion is (4), and 585 therefore the likelihood has the role of an objective function 586 to minimize. This objective function measures the root- 587 mean-square error (RMSE) between calculated and mea- 588 sured heads at the nine observation wells. With such loose 589 constraints, we ensure that the problem is severely ill posed 590 and that the solution space has multiple local minima. 591

[50] Six series of runs are performed. The first four series 592 (ISR) use fixed  $\varphi$  values of (a) 0.1, (b) 0.05, (c) 0.01, and 593 (d) 0.005. The last two series (ISR-SA) use a varying 594 sampled fraction  $\varphi$  according to (5). For each series, 25 595 optimizations are performed with ISR; each optimization 596 is carried out for  $i = 200$  iterations. Figure 6 displays 597 the evolution of each optimization (thin black line) and 598 the median of each series (bold red line). The parameters 599 of the simulated annealing cooling schedule are (Figure 6e) 600  $c = 0.990$  and (Figure 6f)  $c = 0.995$ , with  $\varphi_{\max} = 0.1$  for 601 both series. Figure 6g shows the evolution of  $\varphi$  as a function 602 of the iterations for both cooling schemes, with  $c = 0.990$  603 representing a fast cooling and  $c = 0.995$  representing a 604 slower cooling. Table 1 provides a summary of the RMSE 605 values obtained with each series of runs. Median values 606 illustrate the overall performance for each series and mini- 607 mum/maximum values help identify the stability of the 608 optimization behavior. 609

[51] Nearly all parameters are, on average, able to reduce 610 the RMSE by more than 1 order of magnitude in 200 611 iterations. The only exception is  $\varphi = 0.1$ , whose poor per- 612 formance can be explained by too small steps between one 613 solution and another (large fraction of resampled nodes). 614 Although this does not prevent reaching a local minimum, it 615 can significantly reduce the convergence rate if the topog- 616 raphy of the objective function is very flat. All other  $\varphi$  617 values tested achieve similar median fits. Constant resam- 618 pling with  $\varphi = 0.05$  performs slightly better for the median 619 and the best fit than smaller  $\varphi$  values, but this comes at the 620 price of having some optimizations that did not converge, 621



**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.

t2.1 **Table 2.** Test Case Using ISR With Direct Sampling<sup>a</sup>

		min RMSE	median RMSE	max RMSE	number of evaluations
t2.2	prior	0.0435	0.9977	3.5924	100,000
t2.3	rejection	0.0435	0.0754	0.1155	100,000
t2.4	Metropolis	0.0486	0.0873	0.1294	26,753
t2.5	interrupted MC	0.0305	0.0745	0.1098	8108

t2.6 <sup>a</sup>Ranges of RMSE and number of forward problem evaluations for the ensembles sampled with different methods.

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

624 [52] The same phenomenon occurs with ISR-SA. If the  
 625 large steps at the beginning of the optimization do not yield  
 626 models close to a good fit, the smaller steps that occur  
 627 later can only provide limited improvements. The search  
 628 then remains away from good areas, and the corresponding  
 629 solutions show poor fits, even after a large number of  
 630 iterations. Conversely, if a suboptimal solution is reached in  
 631 the initial phase of the optimization, the smaller steps that  
 632 occur later allow fine adjustment. This dual and unstable  
 633 behavior explains the presence of both the best and worst  
 634 fits of all series when ISR-SA is used with a quick cooling  
 635 schedule ( $c = 0.990$ ). With a slower cooling ( $c = 0.995$ ), it  
 636 is less pronounced. Due to this high sensitivity to the initial  
 637 sample, a possible strategy could be to choose the starting  
 638 point of the optimization as the best of a small set of ran-  
 639 domly generated realizations.

640 [53] ISR-SA has the potential of achieving better fits, but  
 641 it provides only slightly better median convergence. The  
 642 price to pay is a high sensitivity to the cooling parameter  
 643 that may be difficult to adjust in practice. Conversely, ISR  
 644 does not require the adjustment of cooling parameters, and  
 645 we think this is a major advantage from a practical point of  
 646 view. We tested a large array of  $\varphi$  values, with a factor 20  
 647 between the lowest and highest  $\varphi$  values. Corresponding  
 648 RMSE values vary only with a factor 2; therefore, it seems  
 649 that ISR is not very sensitive to the parameter  $\varphi$ , at least for  
 650 the present case. This is fortunate because it eases the  
 651 adjustment of optimization parameters.

## 652 2.7. Approximating the Posterior with Multiple 653 Optimizations

654 [54] Consider  $n$  independent Markov chains, each using  
 655 acceptance criterion (4) to define the models that are  
 656 accepted in the chain. Taking one optimized model per  
 657 chain yields an ensemble of  $n$  samples, all of them having  
 658 a nonzero prior probability and fitting the data. However,  
 659 Bayes' rule may not have been observed since models are  
 660 sampled from a subset of the prior that may not reflect the  
 661 exact posterior. Samples can belong to the posterior (in the  
 662 sense that they match the data well), but they are not  
 663 necessarily distributed according to the posterior. There-  
 664 fore, a bias is introduced on the modeling of uncertainty  
 665 (here we use the term "bias" in the sense of a faulty  
 666 sampling design).

[55] Such a procedure is a form of importance sampling. 667 The central idea of importance sampling is that certain areas 668 of the prior distribution have more impact on the posterior 669 than others. Hence it may be preferable to avoid proposing 670 samples in regions of low fit (see Smith [1997] for a 671 comprehensive review). Instead of uniformly sampling from 672  $f(\mathbf{m})$ , one wishes to sample models from a biased dis- 673 tribution  $f^*(\mathbf{m})$  that excludes areas of low fit. As a result, 674 sampling is not as imposed by Bayes' rule, but according 675 to a biased posterior. Importance sampling techniques pro- 676 vide an approximate compensation for such bias by intro- 677 ducing a weighting term in the probability of acceptance 678 of a model  $\mathbf{m}$ , weights being given by the ratio of the priors 679  $f(\mathbf{m})/f^*(\mathbf{m})$ . 680

[56] Since importance sampling can greatly accelerate the 681 sampling process, its use in the context of the hydro- 682 geological inverse problem is appealing. However, applying 683 the bias correction in practical cases is problematic because 684 the ratio of priors is difficult to define. Without bias cor- 685 rection, there is no guarantee that samples obtained by 686 multiple optimizations are even approximately distributed 687 according to Bayes' rule. The distribution of the sampled 688 models is dependent on the stopping criterion of the opti- 689 mization process. If the number of iterations is too large, all 690 optimizations converge to the global minimum. In addition 691 to wasting CPU time, it results in an uncertainty smaller than 692 desired. If the number of iterations is too small, a very large 693 portion of the prior is sampled, yielding too high uncer- 694 tainty. In other words, deterministic stopping criteria give 695 little control of whether the data are over- or underfitted. We 696 illustrate this problem with the example of variance mini- 697 mization (Figure 2). We perform 243 optimizations using 698 acceptance criterion (4), and we set as a stopping criterion a 699 total of  $i_{\max} = 10$  accepted models. The resulting ensemble 700  $\mathbf{M}_i^{10}$  (Figure 2a, blue line) presents a distribution of variance 701 much narrower than what is found with rejection sampling 702 (Figure 2a, black solid line). Moreover, the large amount of 703 forward model evaluations (35,927) represents a waste of 704 computer resources. 705

[57] In the case of ISR using acceptance criterion (4), the 706 models in the Markov chain are drawn from the biased prior 707  $f^*(\mathbf{m}) = f(\mathbf{m}|\mathbf{m}_i)$ , which is the ensemble of all realizations 708 obtained by extracting a subset  $\mathbf{r}_i$  from the previous member 709 of the chain  $\mathbf{m}_i$ . At each iteration,  $f^*(\mathbf{m})$  is more biased 710 toward high fits but still remains a subset of the prior, as 711 shown in section 2.3. Therefore, a first bias is that the 712 likelihood of the models is too high. 713

**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).

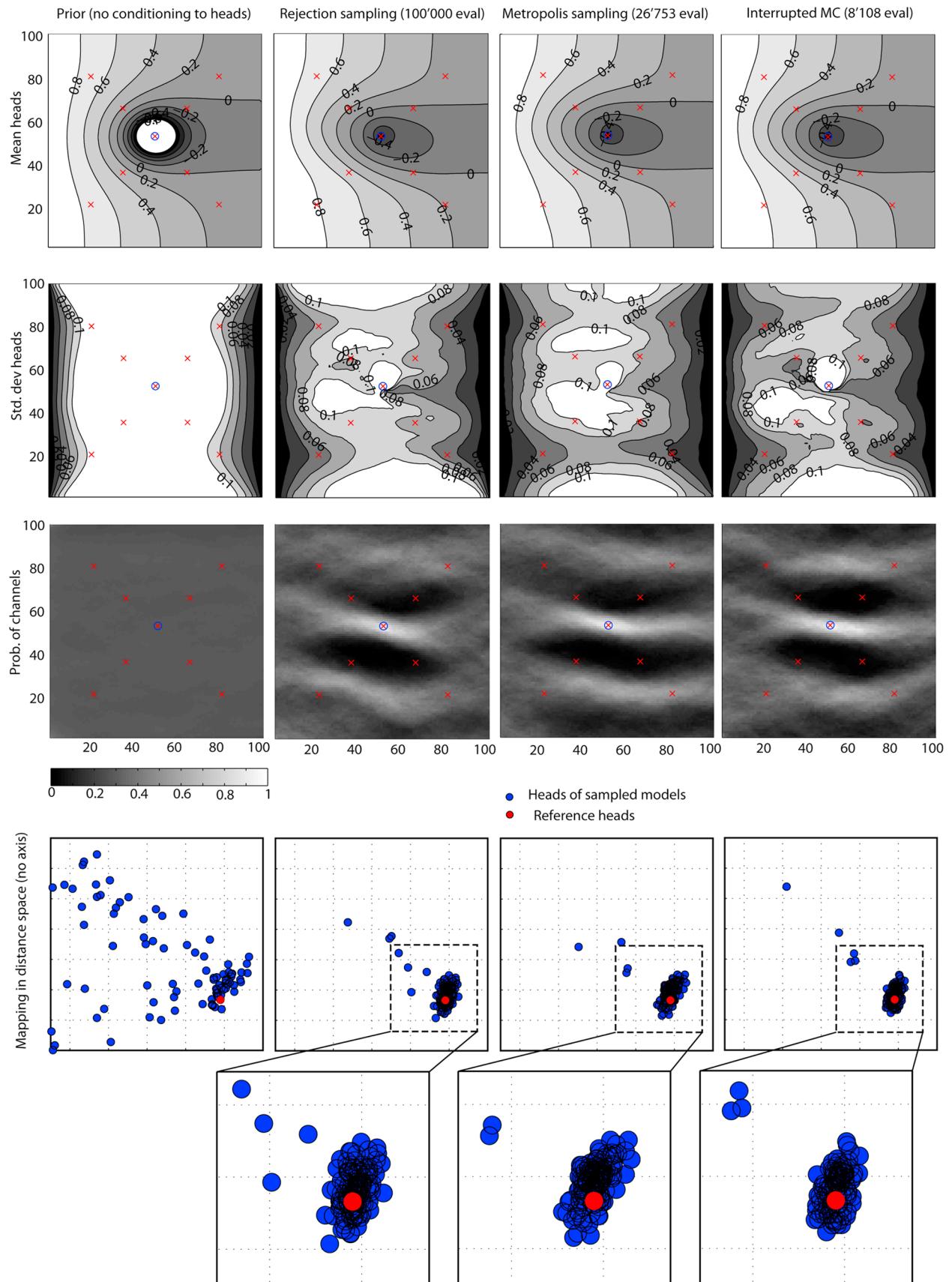


Figure 9

[58] As a practical bias correction, we propose to 714 maturely interrupt and sample the chain, with a criterion 715 based on the likelihood. Our idea relies on the fact that for a 716 proposal model  $\mathbf{m}^*$  to be submitted to the possibility of 717 acceptance, all previous models in the chain  $\mathbf{m}_1, \dots, \mathbf{m}_i$  718 must also have been submitted to this same possibility 719 and rejected. In other words, the existence of a model is 720 conditioned to the rejection of all of its predecessors. 721 Hence, the probability that a model is even considered as 722 a sample decreases with the iterations, which is a second 723 bias on  $f^*(\mathbf{m})$ , but in the opposite direction. Models are 724 increasingly likely to be accepted, but they are less and 725 less likely to be submitted to the acceptance criterion. 726 Although they are difficult to define precisely, both effects 727 are opposite and may compensate each other.

[59] To obtain one sample by interrupted Markov chain, 729 one needs to design an ever-improving Markov chain that 730 accepts new members under condition (4). The chain should 731 be interrupted following a stochastic stopping criterion 732 similar to the acceptance rule of rejection sampling. This is 733 can be accomplished in the following steps:

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

[61] 2. Generate an initial model  $\mathbf{m}_1$  using a geostatistical 736 simulation algorithm, and evaluate its likelihood  $L(\mathbf{m}_1)$ .

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

[63] a. Select randomly a subset  $\{\mathbf{r}_i = Z_i(\mathbf{x}_\alpha), \alpha = 1, \dots, n\}$  739 of  $n$  points belonging to  $\mathbf{m}_i$ .

[64] b. Generate a proposal realization  $\mathbf{m}^*$  by conditional 741 simulation using  $\mathbf{r}_i$  and the same geostatistical model with a 742 new random seed.

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

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

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

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

[69] iii. If  $u \leq P(\mathbf{m}^*)$ , accept  $\mathbf{m}^*$  as a sample of the pos- 748 terior distribution and interrupt the chain. If  $u > P(\mathbf{m}^*)$ , 749 continue the chain.

[70] e. Attempt to narrow down  $f^*(\mathbf{m})$ : if  $L(\mathbf{m}^*) \geq L(\mathbf{m}_i)$ , 751 set  $\mathbf{m}_{i+1} = \mathbf{m}^*$ , otherwise go back to a (i.e., do not 752 increment  $i$ ).

[71] The algorithm above is a form of rejection sampler 754 that samples from the proposal models of a chain instead of 755 uniformly sampling the prior. It is indeed a biased sampler 756 compared to a rejection sampler (except for the first iter- 757 ation, where it is exactly a rejection sampler). It is a heuristic 758 way to quickly obtain an approximation of the posterior, 759 and it does not replace exact samplers. However, it still 760 accounts for the likelihood function, which is not the case 761 with deterministic stopping criteria such as a fixed num- 762 ber of iterations, a maximum number of iterations without 763 improvement, or a threshold in the objective function. 764 More importantly, interrupting the chains reduces com- 765 putational burden by skipping the unnecessary runs that 766 incur overfitting. Incidentally, since each Markov chain is 767 independent, the approach is straightforward to parallelize 768 [Mariethoz, 2010].

[72] As a preliminary test of the interrupted Markov 770 chains, we use it to evaluate the posterior distribution of 771 the minimum variance problem (Figure 2). We generate 772 243 samples with interrupted Markov chains, using like- 773 lihood (3) and the same supremum value that was used for 774 rejection sampling. The variance distribution of the resulting 775 ensemble  $\mathbf{M}_i$  is displayed in Figure 2a (green line), and the 776

variograms of all models are shown in Figure 2e. As 777 expected,  $\mathbf{M}_i$  does not display exactly the same distribution 778 of variance and variograms as the ensembles obtained with 779 rejection and Metropolis samplers. However, the bias is less 780 than with a fixed number of 10 iterations ( $\mathbf{M}_i^{10}$ ), and it 781 requires much less model evaluations. The algorithm of 782 interrupted Markov chains is able to obtain a reasonable 783 ensemble with only 3871 evaluations, whereas rejection and 784 Metropolis sampling need 20,000 evaluations and a fixed 785 number of 10 iterations requires 35,927 evaluations. 786

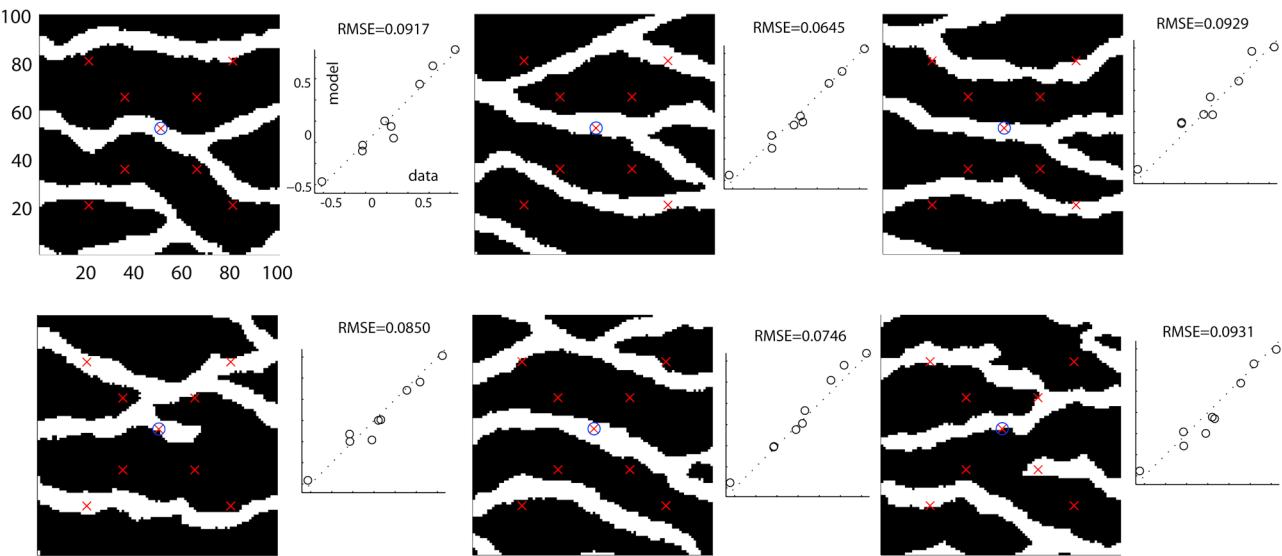
### 3. Test Case

#### 3.1. Problem Setting

[73] One of the key features of ISR is that its principle is 789 not associated with a specific simulation method or a certain 790 type of spatial variability. In the preceding section, we 791 presented ISR with multi-Gaussian examples. To demon- 792 strate the general applicability of ISR, we define a new 793 problem involving sand channels in a clay matrix and we 794 use the direct sampling method (DS) to model it [Mariethoz 795 and Renard, 2010; Mariethoz et al., 2010]. This technique 796 uses multiple-point statistics, which are well-suited to 797 model a wide range of structural models, multi-Gaussian or 798 not [Caers, 2003, 2005; Guardiano and Srivastava, 1993; 799 Hu and Chugunova, 2008; Journel and Zhang, 2006; 800 Strebelle, 2002]. DS has the particularity that it does not 801 determine the values of the simulated nodes by drawing 802 them from local probability distributions. Instead, values are 803 directly sampled from a training image. Therefore, inverse 804 modeling methods that rely on the perturbation of the local 805 distribution, such as GDM and PPM, cannot be applied. 806 Since DS allows conditioning to points data, we show that 807 ISR can be applied. 808

[74] The spatial model of the sand/clay patterns is defined 809 by the categorical training image displayed in Figure 7, 810 representing sand channels in a clay matrix [Strebelle, 2002]. 811 With this training image and the parameters described 812 below, one realization is generated on a grid of 100 by 813 100 nodes, which is thereafter considered as the reference 814 field (Figure 8a). Parameters of the simulation are a neigh- 815 borhood of  $n = 25$  nodes and a distance threshold set to 816  $t = 0.04$ . The meaning of these parameters is that, for any 817 simulated node, the data event (pattern) made of the 25 closest 818 neighbors is considered. Starting from a random location, 819 the training image is scanned until encountering a node 820 whose neighborhood matches at least 24 out of the 25 nodes 821 searched for. (The parameter  $t$  represents the fraction of 822 mismatching nodes allowed, which here equals 1 since 823  $0.04 \times 25 = 1$ . Hence up to one mismatching node is 824 allowed.) The value of this node is then assigned to the 825 simulated location. The method reproduces the statistics 826 of the training image up to the  $n$ th order [Shannon, 1948]. 827

[75] Although multiple-point algorithms guarantee that 828 conditioning data are locally honored, there may be artifacts 829 in the neighborhood of the conditioning data (H. Kjønsberg 830 and O. Kolbjørnse, Markov mesh simulations with data 831 conditioning through indicator kriging, paper presented at 832 Geostats 2008, 1–5 December 2008). Conversely, kriging 833 offers perfect conditioning at the data locations and in the 834 spatial relationships between conditioning data and sur- 835 rounding locations.. In the case of DS, when a data con- 836 figuration observed in the simulation is not found in the 837



**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.

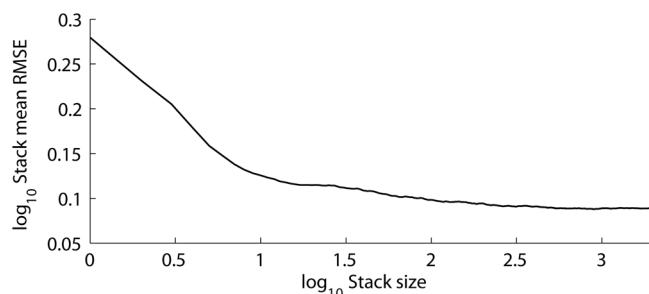
838 training image, DS selects the best matching configuration  
 839 of the training image. In such cases, patterns that are  
 840 incompatible with the prior can occur in the simulation.  
 841 If these patterns are in the neighborhood of data artifacts  
 842 may appear, especially when large amounts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

852 [76] A uniform hydraulic conductivity value of  $10^{-2}$  m/s  
 853 is assigned to sand channels (Figure 7, white) and a value of  
 854  $10^{-4}$  m/s to clays (Figure 7, black). The resulting hydraulic  
 855 conductivity field is used in the same setting as the example  
 856 of the preceding section (Figure 5: one pumping well and  
 857 nine observation wells, and the same boundary conditions).  
 858 The resulting reference heads are displayed in Figure 8b.  
 859 The head is known at the nine observation wells, and the  
 860 RMSE of the calculated versus observed head is considered  
 861 to evaluate a given solution.

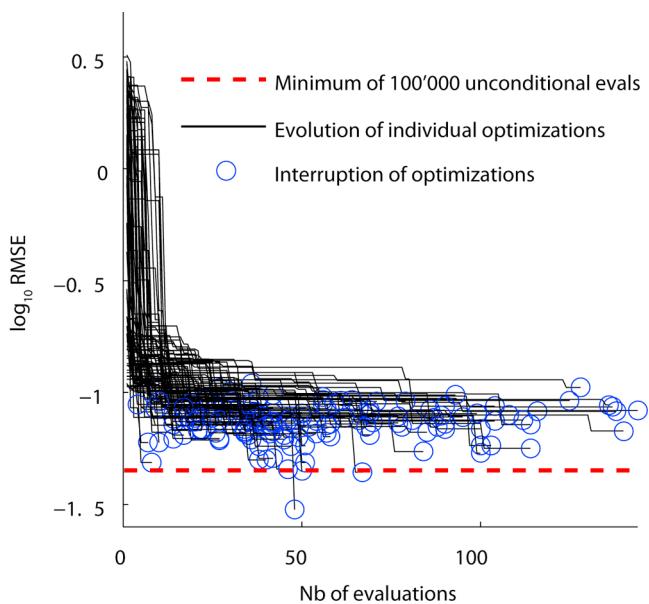
### 862 3.2. Ensemble Solutions with Different Samplers

863 [77] The posterior distribution is characterized using  
 864 different techniques. Table 2 provides the RMSE of the  
 865 calculated heads and the number of forward problem eval-  
 866 uations for each sampling method. Figure 9 summarizes  
 867 the results graphically. Each column represents a sampling  
 868 method, and each row represents a different representation  
 869 of the ensembles of models considered. The first row is the  
 870 ensemble mean head, the second row is the head standard  
 871 deviation, and the third row is the probability of occurrence  
 872 of channels.

[78] 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.

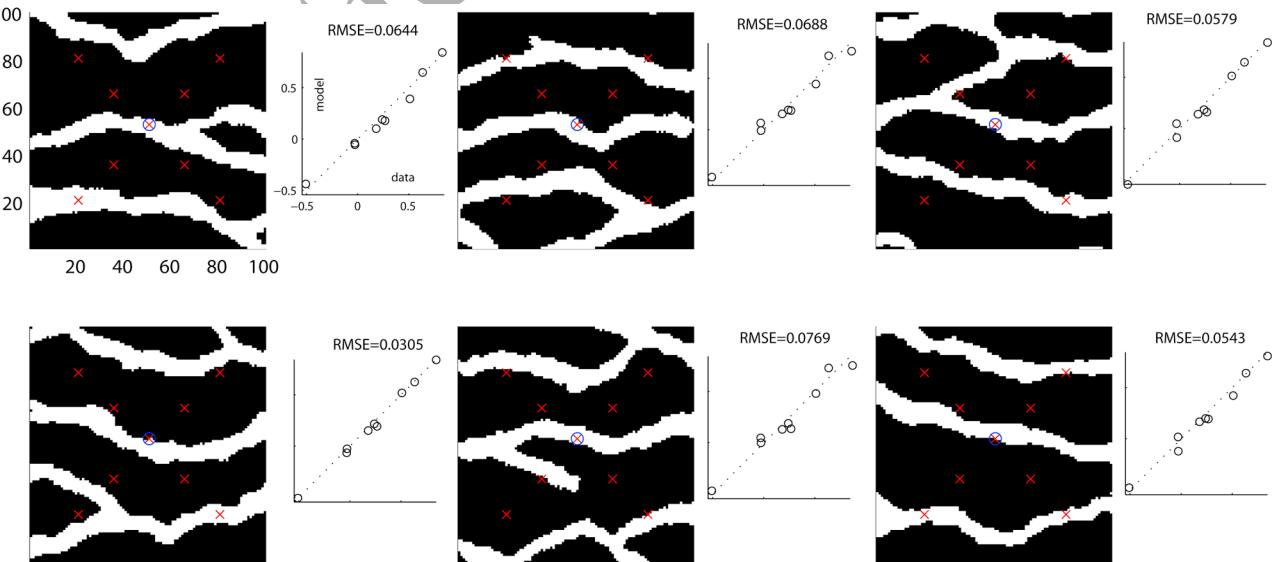


**Figure 11.** Convergence of the Metropolis sampler in the synthetic test case.



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

[79] 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  $-1 \text{ m}$ ). 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.



**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.

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

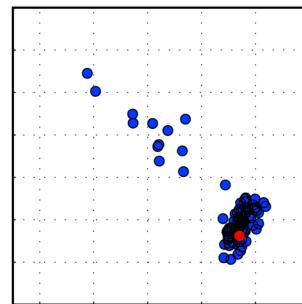
$$L(\mathbf{m}_i) = \exp\left(-\frac{\text{RMSE}(\mathbf{m}_i)^2}{2\sigma^2}\right), \quad (6)$$

with  $\sigma = 0.03 \text{ m}$ , 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  $f(\mathbf{m}|\mathbf{d})$ . 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 non-uniqueness of the solutions and of multiple local minima in the solution space.

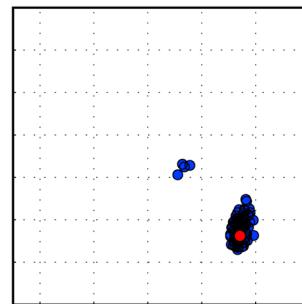
[81] Compared to the prior models, the standard deviation of heads displays reduced uncertainty, especially at the data locations. The probability of occurrence of sand shows that the head measurements captured some essential features governing flow behavior. One such feature is the presence of a channel 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.

[82] 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,

a) Optimizations with 15 iterations



b) Optimizations with 150 iterations



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

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

947 [83] The third column of Figure 9 shows that these  
948 150 samples are similar to the outcomes of rejection sam-  
949 pling. The mean heads and the probability of occurrence of  
950 channels are fairly close to the ones obtained by rejection.  
951 Slight differences are observed for the standard deviation of  
952 heads. In the distance-based representation, both rejection  
953 and Metropolis samplers produce models that are represented  
954 in the distance space as a main cluster with a few outliers.  
955 While the main cluster is similar for both samplers, rejection  
956 produced seven outliers and Metropolis produced only four.  
957 Moreover, Metropolis sampling results in a higher median  
958 RMSE. Although both samplings are fairly similar, the dif-  
959 ferences can be attributed to the relatively small number of  
960 samples (150), but also to the imperfect conditioning of the  
961 DS simulation method.

962 [84] The fourth column of Figure 9 represents 150 sam-  
963 ples obtained by interrupted Markov chains, with a constant  
964 fraction of resampled nodes of  $\varphi = 0.01$ . Likelihood (6) is  
965 used, and the supremum is the same as for rejection sam-  
966 pling. The results are relatively similar to the ensemble  
967 obtained by rejection sampling, using only 8108 forward  
968 problem evaluations (about 54 forward simulation runs for  
969 each matched model). The head standard deviation is  
970 noticeably reduced in the upper part of the image. In the  
971 distance-based representation, five models lie out of the  
972 main cluster, which is similar to what was observed with  
973 rejection sampling. However, the main cluster is too narrow  
974 (see zoomed-in part). Figure 12 shows the evolution of the  
975 150 optimizations and their interruptions. The number of  
976 iterations  $i$  before interruption ranges between 4 and 144,  
977 with an average of 54 iterations. Six optimized realiza-  
978 tions obtained by interrupted Markov chains are shown in  
979 Figure 13. Similarly to the case of rejection sampling, the  
980 presence of diversity in the population of solutions indicates  
981 that different local minima have been explored.

982 [85] For comparison, Figure 14 displays the distance-  
983 based representation of models obtained with deterministic  
984 stopping criteria, after a fixed number of iterations  $i_{\max} =$   
985 150 and  $i_{\max} = 15$ . Clearly, 15 iterations are not enough and  
986 produce an ensemble that is too spread, while 150 iterations

are too much, only representing a narrow subset of the  
987 desired posterior. Note that the correct number of iterations  
988 cannot be known a priori.  
989

#### 4. Conclusion

[86] We presented the iterative spatial resampling method  
991 (ISR) to perturb realizations of a spatially dependent vari-  
992 able while preserving its spatial structure. The method is  
993 used as a transition kernel to produce Markov chains of  
994 geostatistical realizations. Depending on the acceptance/  
995 rejection criterion in the Markov process, it is possible to  
996 obtain a chain of realizations aimed either at characterizing a  
997 certain posterior distribution with Metropolis sampling or at  
998 calibrating one realization at a time. ISR can therefore be  
999 applied in the context of Bayesian inversion or as an optimi-  
1000 zation method. For the latter case, we present a stopping  
1001 criterion for optimizations inspired from importance sam-  
1002 pling. In the studied cases, it yields posterior distributions  
1003 reasonably close to the ones obtained by rejection sampling,  
1004 with important reduction in CPU cost.  
1005

[87] The method is based solely on conditioning data;  
1006 hence it can be used with any geostatistical technique  
1007 able to produce conditional simulations. Moreover, ISR can  
1008 be straightforwardly implemented without modification of  
1009 existing computer codes. The method is simple in its con-  
1010 cept and needs very little parameterization.  
1011

[88] The fraction of resampled nodes  $\varphi$  is the only  
1012 parameter required for optimization with ISR. It has been  
1013 shown that the method is efficient for a wide range of  $\varphi$ . This  
1014 low sensitivity is a major advantage from a practical point of  
1015 view because it saves the user the hassle of performing  
1016 lengthy sensitivity analysis to find optimal parameters.  
1017

[89] The approach is illustrated with both continuous and  
1018 discrete variables. We use head data and groundwater flow  
1019 problems, but the principle is general and can be applied  
1020 to other inversion problems such as the ones involving  
1021 geophysical applications. Future research will focus on  
1022 extending the concept of ISR. For example, local perturba-  
1023 tions can be obtained by resampling certain areas more than  
1024 others or by using quasirandom resampling [e.g., Tang et  
1025 al., 2008]. This could be used when the forward problem  
1026 provides local fitness or sensitivity information. Another  
1027 aspect is the integration of preferential search directions. In  
1028 this paper, we investigated search patterns that use random  
1029 search directions, obtained by sampled locations that are not  
1030 correlated between an iteration and the next one. It may be  
1031

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.

[90] **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. Zuccolo (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. Kopperberg (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 V. e. a., pp. 831–849, D. Reidel.
- De Marsily, G., F. Delay, J. Gonçalves, 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 inverting, *Water Resour. Res.*, **31**(10), 2411–2419.
- 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 non-linear 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.*, **46**(10), 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.

- 1178 Omre, H., and H. Tjelmeland (1996), *Petroleum Geostatistics*, Dept. of  
1179 Math. Sci., Norwegian Univ. of Sci. and Technol., Trondheim, Norway.  
1180 Pan, L., and L. Wu (1998), A hybrid global optimization method for  
1181 inverse estimation of hydraulic parameters: Annealing-simplex method,  
1182 *Water Resour. Res.*, 34(9), 2261–2269.  
1183 RamaRao, B., A. LaVenue, G. De Marsily, and M. Marietta (1995), Pilot  
1184 point methodology for automated calibration of an ensemble of condi-  
1185 tionally simulated transmissivity fields. 1. Theory and computational  
1186 experiments, *Water Resour. Res.*, 31(3), 475–493.  
1187 Remy, N., A. Boucher, and J. Wu (2009), *Applied Geostatistics with  
1188 SGeMS: A User's Guide*, 284 pp., Cambridge Univ. Press, Cambridge,  
1189 U. K.  
1190 Ronayne, M., S. Gorelick, and J. Caers (2008), Identifying discrete geo-  
1191 logic structures that produce anomalous hydraulic response: An inverse  
1192 modeling approach, *Water Resour. Res.*, 44, W08426, doi:10.1029/  
1193 2007WR006635.  
1194 Scheidt, C., and J. Caers (2009), Representing spatial uncertainty using dis-  
1195 tances and kernels, *Math. Geosci.*, 41(4), 397–419.  
1196 Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst.  
1197 Tech. J.* 27, 379–423.  
1198 Smith, J. (1997), Quick simulation: A review of importance sampling tech-  
1199 niques in communications systems, *IEEE J. Sel. Areas Commun.*, 15(4),  
1200 597–613.  
1201 Strebelle, S. (2002), Conditional simulation of complex geological struc-  
1202 tures using multiple-point statistics, *Math. Geol.*, 34(1), 1–22.  
1203 Subbey, S., M. Christie, and M. Sambridge (2004), Prediction under uncer-  
1204 tainty in reservoir modeling, *J. Pet. Sci. Eng.*, 44(1–2), 143–153.  
1205 Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2008), Comparing  
1206 sensitivity analysis methods to advance lumped watershed model identi-  
fication and evaluation, *Hydrol. Earth Syst. Sci.*, 11(2), 793–817, doi:10.5194/hess-11-793-2007.  
1207 Tarantola, A. (2005), *Inverse Problem Theory and Methods for Parameter  
estimation*, Soc. Ind. Appl. Math., Philadelphia, Pa.  
1208 Tjelmeland, H., and J. Besag (1998), Markov random fields with higher-  
1209 order interactions, *Scand. J. Stat.*, 25, 415–433.  
1210 Vesselinov, V., S. Neuman, and W. Illman (2001), Three-dimensional  
1211 numerical inversion of pneumatic cross-hole tests in unsaturated frac-  
1212 tured tuff. 1. Methodology and borehole effects, *Water Resour. Res.*,  
1213 37(12), 3001–3017.  
1214 von Neumann, J. (1951), Various techniques used in connection with ran-  
1215 dom digits. Monte Carlo methods, *Natl. Bur. Stand.*, 12, 36–38.  
1216 Yeh, W. (1986), Review of parameter identification procedures in groundwa-  
1217 ter hydrology: The inverse problem, *Water Resour. Res.*, 22(2), 95–108.  
1218 Zanini, A., and K. Kitanidis (2009), Geostatistical inversing for large-con-  
1219 trast transmissivity fields, *Stochastic Environ. Res. Risk Assess.*, 23(5),  
1220 656–677.  
1221 Zheng, C., and P. Wang (1996), Parameter structure identification using  
1222 tabu search and simulated annealing, *Adv. Water Resour.*, 19(4),  
1223 215–224.  
1224 Zimmerman, D., et al. (1998), A comparison of seven geostatistically based  
1225 inverse approaches to estimate transmissivities for modeling advective  
1226 transport by groundwater flow, *Water Resour. Res.*, 34(6), 1373–1413.  
1227
- 
- J. Caers, ERE Department, Stanford University, 367 Panama St., Rm. 65,  
1228 Stanford, CA 94305-2220, USA.  
G. Mariethoz and P. Renard, Centre for Hydrogeology, University  
1229 of Neuchâtel, 11 Rue Emile Argand, CP 158, CH-2009 Neuchâtel,  
1230 Switzerland. (gregoire.mariethoz@minds.ch)