History matching with Iterative Spatial Resampling (ISR)

Gregoire Mariethoz, Jef Caers, Philippe Renard

Department of Energy Resources Engineering
Stanford University

Abstract

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 underdetermined problems consist in 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. McMC methods allow characterizing this ensemble, but are difficult to apply due to the computational burden involved. Alternatively, optimization (calibration) methods are more efficient, but do typically provide an ensemble of solution with unrealistically narrow uncertainty and possibly conflicting with the prior model distribution.

Both frameworks (McMC and optimization) rely on a perturbation mechanism to steer the search for solutions. In this context, we propose a general transition kernel (Iterative Spatial Resampling, ISR) that preserves any spatial model produced by conditional simulation.

A realization is perturbed by extracting a subset of it as randomly located points, and by using these points as conditioning data to generate a new, perturbed model. We demonstrate the applicability of ISR in both contexts of Bayesian inverse problem and optimization. The size of the extracted subset is the only parameter required by ISR. It is shown that optimization remains efficient for a wide range of subset sizes.

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 exact samplers (such as rejection sampler), but with comparatively limited 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 to sample from very different priors, and to condition to a variety of data types.

1. Introduction

Inverse problems are a crucial aspect of reservoir modeling, especially in the context of engineering prediction and decision making problems. Such models need 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 is addressed very efficiently by most geostatistical simulation algorithms [*Deutsch and Journel*, 1992; *Remy et al.*, 2009]. In this paper, we refer to conditioning realizations to indirect state data such as pressure of production data.

Problems involving flow in underground media typically present more unknowns than data. For example, modeling of 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. In a Bayesian formulation, one wants to obtain a posterior distribution given a certain prior distribution and a likelihood function. In this respect, only McMC methods have been shown to sample reasonably accurately from this posterior [*Mosegaard and Tarantola*, 1995; *Omre and Tjelmland*, 1996], i.e. to generate model realizations that 1) match the hard and indirect state data, 2) reproduce for each inverse solution the statistics specified in the prior (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 follow these three requirements.

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. 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.

Therefore, depending on the computational burden involved, it may be appropriate to perform either McMC or optimization of one model at a time. The framework we present in this paper allows dealing with both aspects. It is emphasized that the 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.

This paper is organized as follows. The first part introduces the concept of perturbation by Iterative Spatial Resampling (ISR), explores its properties for both Bayesian inversion and optimization, and performs numerical tests. The second part 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

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

$$f(\mathbf{m} \mid \mathbf{d}) = \frac{f(\mathbf{d} \mid \mathbf{m}) f(\mathbf{m})}{f(\mathbf{d})}. \tag{1}$$

In that formulation, the prior distribution $f(\mathbf{m})$ of the possible models $\mathbf{m}$ can be obtained by performing stochastic realizations not conditioned to the state variables $\mathbf{d}$. The likelihood $f(\mathbf{d}|\mathbf{m})$ defines the probabilistic relationship between the observed state variables $\mathbf{d}$ (the data) and the model $\mathbf{m}$. The computation of the likelihood generally requires running a forward model. The expression $\mathbf{d} = g(\mathbf{m})$ denotes this forward problem. We will denote $L(\mathbf{m})$ the likelihood function, that usually represents the misfit between the observed data and the prediction of a model.

Tarantola [2005] gives a comprehensive overview of the available exact methods to obtain representative samples of $f(\mathbf{m}|\mathbf{d})$. Among them, rejection sampling [*von Neumann*, 1951] and Metropolis sampling [*Metropolis et al.*, 1953] are often used. Both methods do not need the density $f(\mathbf{d})$ to be specified. 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 sub-sampling 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

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

where $L(\mathbf{m})_{\max}$ denotes the supremum, which can be any number equal or above the maximum likelihood value that can be taken by $L(\mathbf{m})$. The distribution of the resulting samples follow $f(\mathbf{m}|\mathbf{d})$. Since it requires a large number of evaluations of $g(\mathbf{m})$, the rejection method is inefficient, but will serve as a reference sampler in this paper.

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 to keep only one every m samples, where m should be large enough for the chain to "forget" the previously accepted models.

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 $\mathbf{m}^*$.

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 honors two particular properties. Firstly, it should be symmetric, such that $Q(\mathbf{m}_i|\mathbf{m}^*) = Q(\mathbf{m}^*|\mathbf{m}_i)$. Secondly, it is desirable that the proposal distribution yields models that do not violate the prior. Typical for hydrogeological inverse problems is that the models are spatially dependant variable fields (such as hydraulic conductivity or porosity), often obtained using geostatistical simulation techniques. An appropriate proposal distribution would need to preserve the spatial dependence modeled in the geostatistical method (either variogram-based, Boolean or MPS). The method we show in this paper honors these two properties.

Previous studies have investigated Markov chains applied to spatially dependent variables, using different proposal (or perturbation) mechanisms. Oliver, et al. [1997] create a 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 method (BMcMC) that incurs local perturbations by successively re-simulating a square area of the realizations (a block). BMcMC has been used for sampling the posterior

distribution of synthetic inverse problems in a multiGaussian framework [*Fu and Gomez-Hernandez*, 2009a; *Fu and Gomez-Hernandez*, 2009b].

Optimization methods aim at finding models maximizing 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 corresponding to 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 [e.g. *De Marsily et al.*, 1984] or gradient free methods [e.g. *Bayer et al.*, 2007; *Karpouzos et al.*, 2001; *Singh et al.*, 2008; *Tsai et al.*, 2003]. 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.

Simulated annealing [*Kirkpatrick et al.*, 1983] has been extensively used in groundwater modeling [e.g. *Pan and Wu*, 1998; *Zheng and Wang*, 1996]. Genetic algorithms [*Fraser*, 1957; *Goldberg*, 1989] have been used for hydrogeological inverse problems for identifying structures in hydraulic conductivity fields [e.g. *Bayer et al.*, 2007; *Karpouzos et al.*, 2001; *Singh et al.*, 2008; *Tsai et al.*, 2003]. Alcolea and Renard [*Alcolea and Renard*, in press] apply BMcMC in conjunction with simulated annealing to optimize non-multiGaussian random fields generated by a multiple-point simulation approach.

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; *Llopis-Albert and Cabrera*, 2009; *Llopis-Albert and Capilla*, 2009; *Ronayne et al.*, 2008]. GDM and PPM are limited to methods that draw simulated values from local probability distributions. For example, they cannot be applied to the Direct Sampling method (DS), that does not draw values from a distribution but directly from a training image [*Mariethoz et al.*, submitted] or a training dataset [*Mariethoz and Renard*, 2010].

Since optimization methods essentially search for solutions corresponding to certain constraints, they may wander outside the prior. This is the case for the Gradual Deformation Method (GDM) which can produce solutions that do not honor the prior [*Caers*, 2007; *Le Ravalec-Dupin and Noetinger*, 2002; *Liu and Oliver*, 2004]. The regularized pilot points method (RPPM) 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]. The Probability

Perturbation Method [*Caers and Hoffman*, 2006] and the Blocking Moving Window [*Alcolea and Renard*, in press] approximate well a uniform sampling from the posterior, at least for the examples presented in those papers.

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 using optimization. We show that in both cases, it allows accurately sampling the posterior. To validate the results, we use the rejection sampler as a reference.

2.2. Iterative Spatial Resampling

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 a generating algorithm.

To implement sampling and searching strategies, one needs to create a chain of dependant realizations. Consequently, one does not wants to draw proposal models $\mathbf{m}^*$ from $f(\mathbf{m})$, but from $f(\mathbf{m}|\mathbf{m}_i)$, $\mathbf{m}_i$ being the previous model in the chain. To avoid violating the prior, the conditional term should ideally be incorporated in the method generating unconditional realizations. Since most unconditional simulation methods also allow generating realizations conditioned to hard data, we propose to use this conditioning capability to impose a conditional term on the prior. More specifically, dependence between $\mathbf{m}^*$ and $\mathbf{m}_i$ is introduced by extracting a subset of realization $\mathbf{m}_i$ as randomly located points $\mathbf{r}_i = \{Z_i(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$ 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 $\mathbf{r}_i$, itself a subset of $\mathbf{m}_i$.

Creating a Markov chain using ISR is accomplished by performing the following steps at each iteration i:

1. Generate an initial model $\mathbf{m}_1$ using a geostatistical simulation algorithm, and evaluate its likelihood $L(\mathbf{m}_1)$.
2. Iterate the following steps:
   a. Select $\mathbf{r}_i = \{Z_i(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$ as a subset of $\mathbf{m}_i$.
   b. Generate a proposal realization $\mathbf{m}^*$ by conditional simulation using $\mathbf{r}_i = \{Z_i(\mathbf{x}_\alpha), \alpha = 1, \ldots, n\}$.
   c. Evaluate $L(\mathbf{m}^*)$.

d. Accept or reject $\mathbf{m}^*$. If the acceptation criterion is the one proposed by Mosegaard and Tarantola [1995], the chain is a Metropolis sampler.

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 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 **r**.

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 φ of the total number of nodes in $\mathbf{m}_i$.

It is important to note that the selected data set **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 are samples from the prior $f(\mathbf{m})$. The only requirement is that the conditioning on **r** is correct, or in other words that the conditioning method does not introduce artifacts into the simulation, neither does it artificially reduce uncertainty.

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

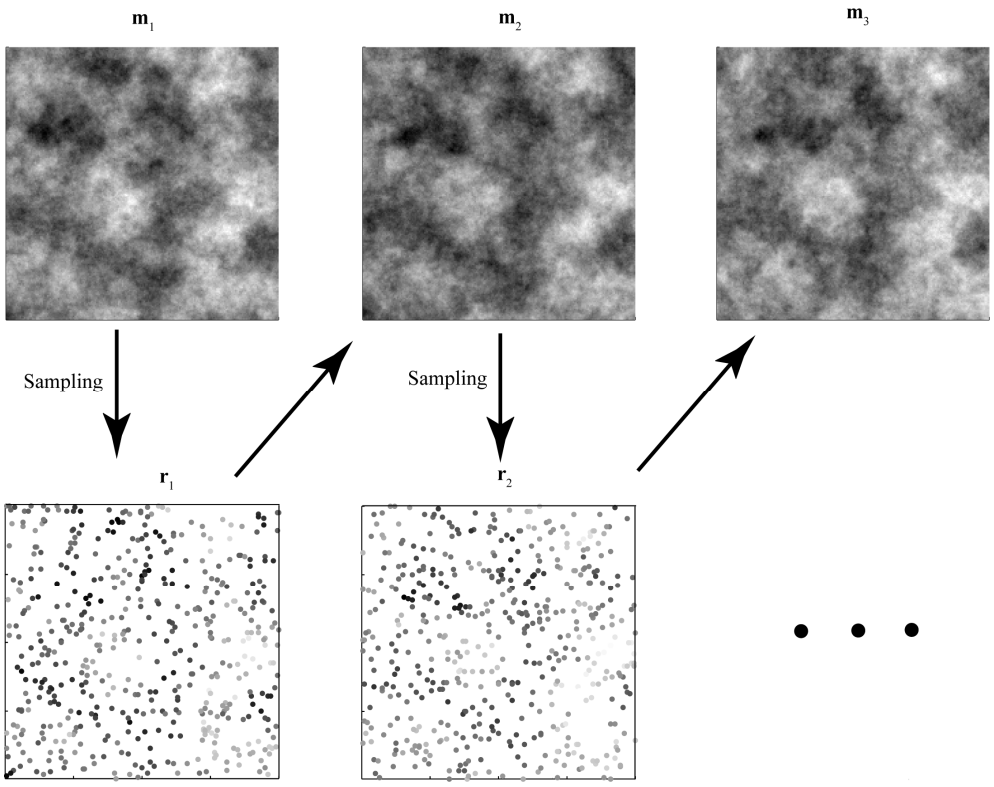

Figure 1. Sketch of the iterative spatial resampling method. An initial realization $\mathbf{m_1}$ is sampled randomly to obtain the subset $\mathbf{r_1}$, which is used as conditioning data for generating another realization $\mathbf{m_2}$. $\mathbf{m_2}$ displays similar local features than $\mathbf{m}_1$ due to the constraints imposed by the conditioning data, but it is also different since the simulation has produced new values at non-conditioned locations.

2.3. Sampling properties of ISR

Several factors may affect the accuracy of the sampling with Metropolis. 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, necessarily involves some degree of approximation. Moreover, convergence of the chain must be reached before performing any sampling, and this may be difficult to assess.

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 could be $\mathbf{m}_1$, $\mathbf{m}_2$, or an entire manifold of models, all of them uniformly distributed. However, since geostatistical simulations are algorithmically defined [*Boucher*, 2007], they may not

offer perfect conditioning, thus making the proposal possibly non-symmetric and incompatible with the prior. 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].

2.4. Using ISR for optimization

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 acceptation criterion (step 2d):

$$\text{if } L(\mathbf{m}^*) \geq L(\mathbf{m}_i), \text{ accept } \mathbf{m}^*. \tag{3}$$

The resulting McMC process is a stochastic search for a single calibrated model. Figure 2 schematically depicts how a simple 2D solution space is explored. The background image represents the real, unknown solution space, with a single global minimum in the center of the image. 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 φ). 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 sub-optimal solutions. Since the search is stochastic, the global minimum will be reached after an infinite number of iterations. However, in most practical applications, it will remain in a local minimum. Because (3) only considers the rank of a proposal solution compared to a previous one, the search is similar to the minimization of an objective function. The likelihood of the final solution depends on algorithmic parameters, such as the stopping criterion used.

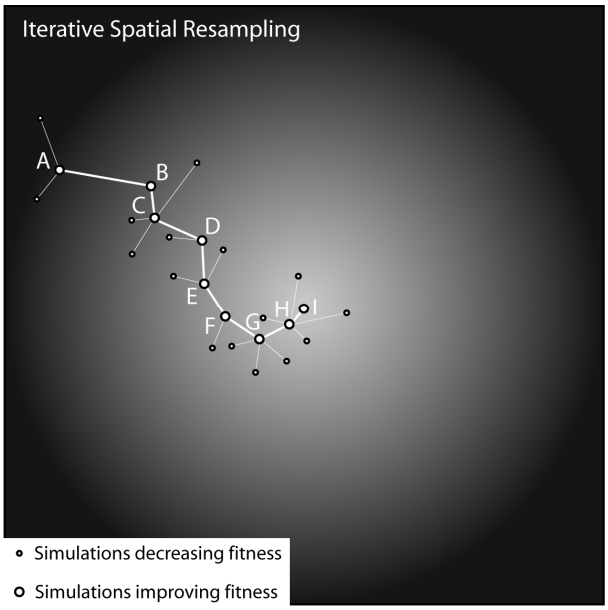

Figure 2. Schematic representation of the search strategy employed by ISR in a simple solution space having 2 degrees of freedom. The background color represents the actual unknown solution space, with values ranging from black (bad solution) to white (good solution). The global minimum is in the center of the domain.

So far a constant fraction of resampled nodes has been considered, but alternatives can be envisioned for optimization. For example, φ can increase at each iteration i. The optimization starts with large steps (exploration phase) and finishes with small steps (narrowing phase). One possible way to accomplish this is with a power law:

$$\varphi(i) = \varphi_{\max}(1 - c^i), \quad 0 \le c \le 1. \tag{4}$$

The larger c, the slower $\varphi(i)$ will reach φ_{max}. This is similar to simulated annealing, with the parameter c defining the cooling schedule and φ_{max} the maximum value of φ (after an infinite number of iterations). Yet, alike other simulated annealing algorithms, adjusting c can be tedious. Nevertheless, we will see below that using (4) can accelerate the convergence compared to keeping φ constant.

2.5. Sensitivity to φ

φ is the only parameter required by ISR. 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 φ.

The problem setting consists in a square aquifer of 100 m by 100 m, discretized in 100 by 100 elements (Figure 3). A reference $\log_{10}$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 $\log_{10}$ hydraulic conductivity is -3, with a variance of 2. The upper and lower sides of the model are no-flow boundaries, a fixed head BC of 1 m is set on the left side, and a fixed head BC of 0 m on the right side. A pumping well extracting 0.003 m^3/s is set in the center of the domain (blue circle), and 9 observation wells are positioned at the locations of the red crosses on Figure 3 (the pumping well is also an observation well). The problem is solved in steady state. The 9 head measurements are the only constraints used to solve the inverse problem. We do not impose any conditioning in order to not over-constrain the prior.

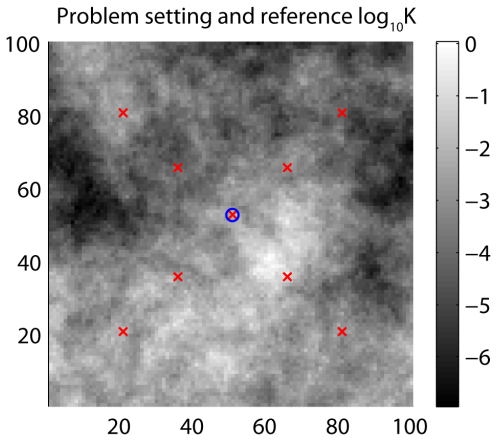

Figure 3: Reference $\log_{10}$ hydraulic conductivity field in m/s. The blue circle marks the location of the pumping well and the red crosses indicate the locations of the observation wells.

Proposal solutions are generated using SGS with the true variogram model. Acceptance criterion is (3), and therefore the likelihood has the role of an objective function to minimize. This objective function consists in the RMSE of the head at the 9 observation wells compared to the heads observed. With such loose constraints, we ensure that the problem is severely ill-posed, and that the solution space has multiple local minima.

For each φ value, a series of 25 optimizations is performed with ISR, each optimization being carried for i = 200 iterations. Figure 4 displays the evolution of each optimization (thin black line) and the median of each series (bold red line). There are six series in total. The first four series (ISR) use a fixed φ 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 φ according to (4). The parameters of the simulated annealing cooling schedule are e) c=0.990 and f) c=0.995, with φ_{max}=0.1 for both series. Figure 4g shows the evolution of φ as a function of the iterations for both cooling schemes, with c=0.990 representing a fast cooling and c=0.995 a slower cooling. Table 1 provides a summary of the RMSE values obtained with each series of runs.

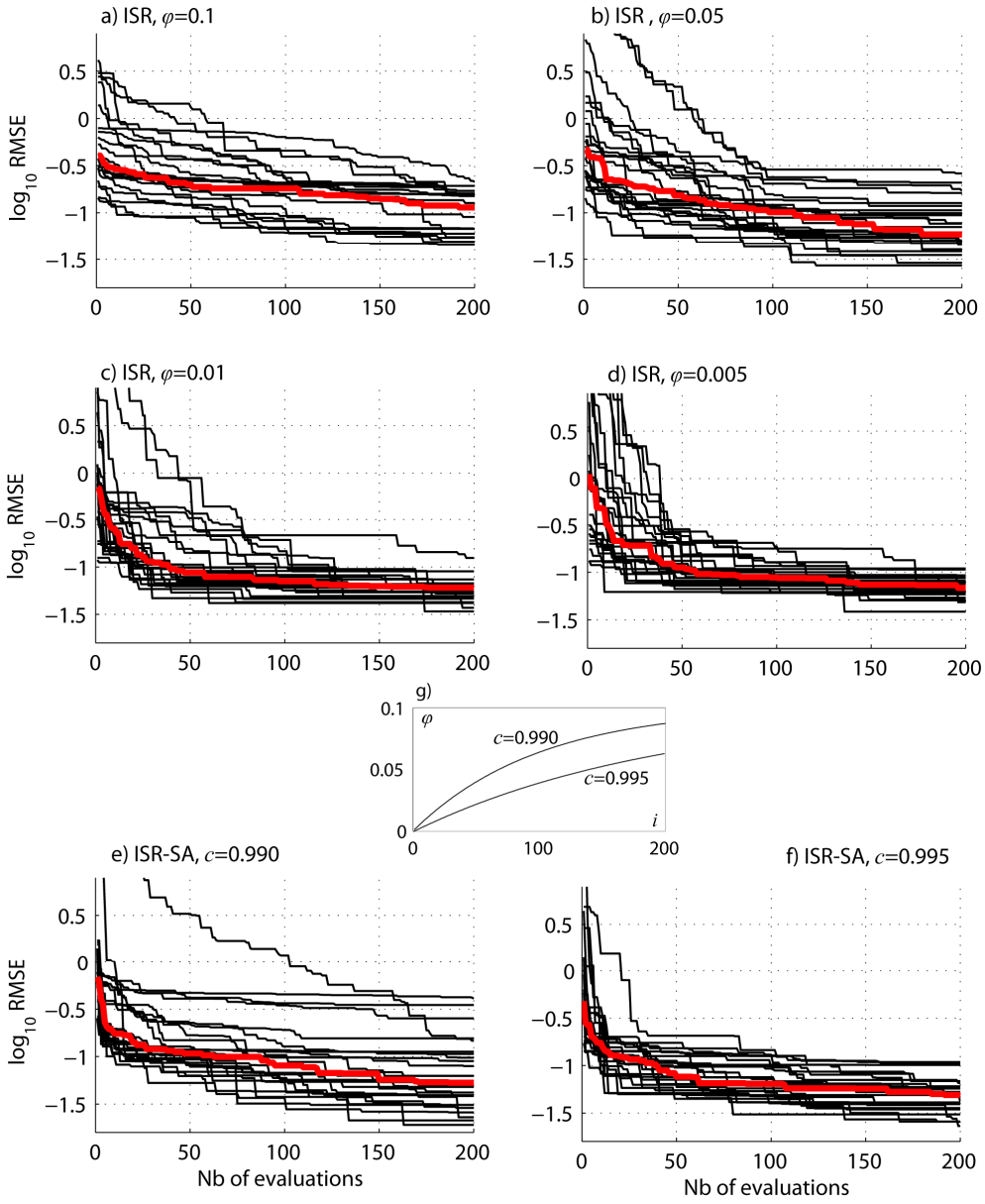

Figure 4. Optimization performance assessment for six different parameterizations of φ. The evolution of each optimization is marked by a thin black line and the median evolution computed from 25 optimizations is shown with a red bold line.

Nearly all parameters are, on average, able to reduce the RMSE of more than one order of magnitude in 200 iterations. The only exception is *φ=0.1,* whose poor

performance can be explained by too small steps between one solution and another (large fraction of resampled nodes). All other φ values tested achieve similar median fits. Constant resampling with φ = 0.05 performs slightly better for the median and the best fit than larger φ values, but this comes at the price of having some optimizations that did not converge, again due to the relatively small steps. This is shown by the maximum RMSE that is larger for φ=0.05.

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 sub-optimal solution is reached in the initial phase of the optimization, the smaller steps that occur later on allow fine adjustment. This dual behavior explains the presence of both the best and the worse 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.

ISR-SA has the potential of achieving better fits, but provides only slightly better median convergence. Conversely, ISR does not require the adjustment of cooling parameters, and we think this is a major advantage from a practical point of view. Since most of the parameters gave good results, it seems that ISR is not very sensible to parameter φ, at least for the present case. This is fortunate because it eases the adjustment of optimization parameters.

Table 1. RMSE of the different optimization runs after 200 iterations.

	minimum RMSE	**median RMSE**	**maximum RMSE**
ISR φ=0.1	0.0451	0.1141	0.2124
ISR φ=0.05	0.0271	0.0591	0.2610
ISR φ=0.01	0.0340	0.0607	0.1242
ISR φ=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

2.6. Approximating the posterior with multiple optimizations

Let n independent Markov chains, each using acceptation criterion (3) 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 belonging to the prior, and also adequately match the data. However, Baye's rule may not have been observed, and models are sampled from a subset of the prior that may not reflect the exact posterior.

Therefore a bias is introduced (here we use the term biased in the sense of a faulty sampling design).

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 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 $f(\mathbf{m}) / f^*(\mathbf{m})$.

Since importance sampling can greatly accelerate the sampling process, using it in the context of 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 correctly distributed. 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 may be sampled, yielding unnecessary high uncertainty. In other words, deterministic stopping criteria give little control on whether the data are over- or under-fitted. Moreover, the large amount of forward model evaluations (35'927) represents a waste of computer resources.

In the case of ISR using acceptation criterion (3), 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 towards high fits, but still remains within the prior, as shown in section 2.3. Therefore a first bias is that the likelihood of the models is too high.

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 $\mathbf{m}^*|\mathbf{m}_i$ 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 less and less likely to be submitted to the acceptance criterion. Although difficult to define precisely, both effects are opposite and may compensate each other.

The following steps yield one sample by interrupted Markov chain:

1. Design an ever-improving Markov chain that accepts new members under condition (8).
2. Define the supremum $L(\mathbf{m})_{max}$.
3. At each iteration i:
   a. Draw a proposal model $\mathbf{m}^*|\mathbf{m}_i$ from $f(\mathbf{m}|\mathbf{m}_i)$ by using $\mathbf{r}_i \subset \mathbf{m}_i$ as conditioning data, and compute the likelihood $L(\mathbf{m}^*)$.
   b. Submit $\mathbf{m}^*|\mathbf{m}_i$ to the possibility of being accepted as a sample of the posterior:
      i. Draw α in $U[0,1]$.
      ii. Compute $P(\mathbf{m}^*|\mathbf{m}_i) = \frac{L(\mathbf{m}^*|\mathbf{m}_i)}{L(\mathbf{m})_{max}}$. If $\alpha < P(\mathbf{m}^*|\mathbf{m}_i)$, accept $\mathbf{m}^*|\mathbf{m}_i$ as a sample of the posterior distribution and interrupt the chain.
   c. If condition (8) is met, set $\mathbf{m}_{i+1} = \mathbf{m}^*|\mathbf{m}_i$, otherwise *go back* to a (i.e. do not increment i).

The algorithm above is 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*, in press].

3. Test case

3.1. Problem setting

One of the key features of ISR is that its principle is not associated to a specific simulation method or a certain kind of spatial model. In the preceding section, we presented ISR with multi-Gaussian examples. To demonstrate the general applicability of ISR, we define a problem involving sand channels in a clay matrix, and we use the simulation method of Direct Sampling (DS) to model it [*Mariethoz et al.*, submitted]. This technique uses multiple-points statistics, which are appropriate to model a wide range of structural models, both multi-Gaussian and non-multi-

Gaussian [*Caers*, 2003; 2005; *Guardiano and Srivastava*, 1993; *Journel and Zhang*, 2006; *Strebelle*, 2002].

The spatial model of the sand/clay patterns is defined by the categorical training image shown in Figure 5, representing sand channels in a clay matrix. With this training image for model, one realization is generated on a grid of 100 by 100 nodes, which is thereafter considered as the reference field (Figure 6a). Parameters of the simulation are a neighborhood of n=25 nodes and a distance threshold of t=0.07. The meaning of these parameters is that for any simulated pixel, the data event (pattern) formed by the 25 closest neighbors is considered. Starting from a random location, the training image is scanned until encountering a pixel whose neighborhood matches at least 24 out the 25 nodes searched for. The value of this pixel is then assigned to the simulated location. The method reproduces the statistics of the training image up to the n^{th} order [*Shannon*, 1948].

Contrarily to kriging, MPS algorithms do not provide perfect conditioning to data [*Kjønsberg and Kolbjørnsen*, 2008]. In the case of DS, when a data configuration observed in the simulation is not found in the training 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, conditioning can be inaccurate, especially when large amounts of conditioning data are present, which is the case with ISR. Consistency between all patterns could be enforced using syn-processing [*Mariethoz et al.*, submitted], that recursively un-simulates and re-simulates nodes until all of them are compatible, but the method has a steep CPU cost. Instead, we use a specific distance between data events that gives a larger relative weight to the nodes corresponding to data. However, perfect conditioning is not guaranteed.

A uniform hydraulic conductivity value of 10^{-2} m/s is assigned to sand channels (white) and a value of 10^{-4} m/s to clays (black). The resulting hydraulic conductivity field is used in the same setting as the example of the preceding section (Figure 3: one pumping well and nine observation wells, and the same boundary conditions). The resulting reference heads are displayed in Figure 6b. Head is known at the 9 observation wells and the RMSE of the modeled versus observed head is considered to evaluate a given solution.

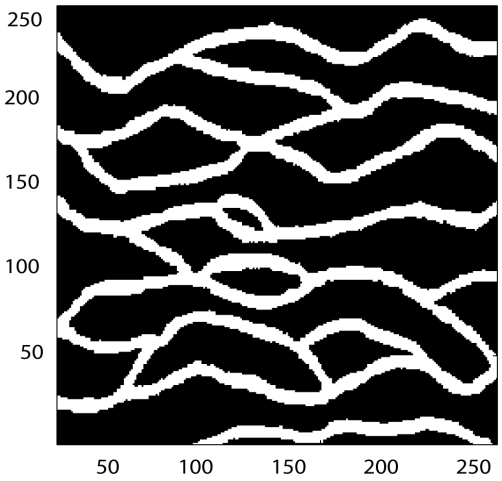

Figure 5. The training image used to model for the sand and clay spatial distribution.

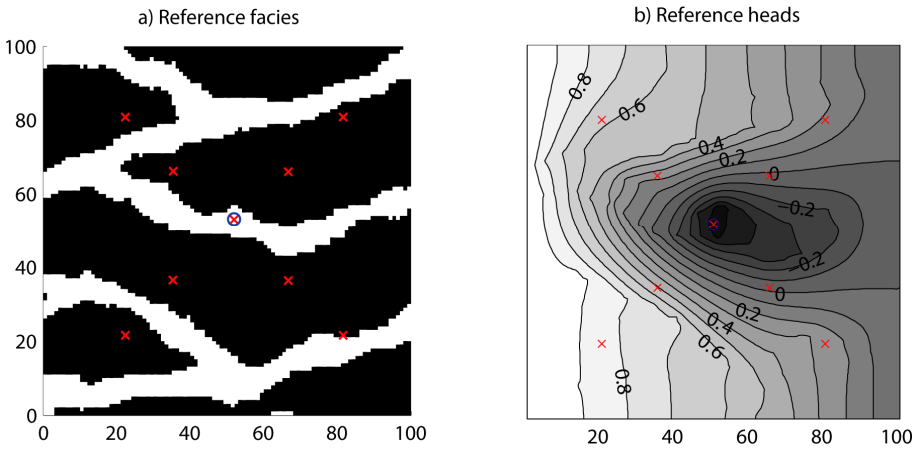

Figure 6: Reference field used for the synthetic test case. The blue circle marks the location of the pumping well and the red crosses indicate of the observation wells. a) Reference facies. b) Corresponding reference heads.

3.2. Ensemble solutions with different samplers

The problem's posterior distribution is characterized with different techniques. **Table 2** provides the RMSE of the modeled heads and the number of forward problem evaluations for each sampling method. Figure 7 summarizes the results graphically. Each column represents a sampling method, and each row a different representation of the ensembles of models considered. The first row is the mean head, the second row its standard deviation, and the third row is the probability of occurrence of channels.

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 **D**, such a representation displays an ensemble of models $\mathbf{m}_i$ as a set of points in a possibly high-dimensional space, arranged in such a way that their respective distances are preserved as much as possible (in a least-squares sense). **D** can be computed using any appropriate measure of distance. For representation, the dimensions carrying less information are often ignored. In the present case, the distance between any two models $d\{\mathbf{m}_i,\mathbf{m}_j\}$ is the pixel-wise Euclidean distance between the heads calculated on the entire domain using both models. **D** is computed using 601 models (150 models for each of the 4 sampling methods, plus the reference, represented by a red dot), but 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.

Table 2. Ranges of RMSE and number of forward problem evaluations for the ensembles sampled with different methods.

	min. RMSE	**median RMSE**	**max. RMSE**	**nb. eval.**
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	8’108

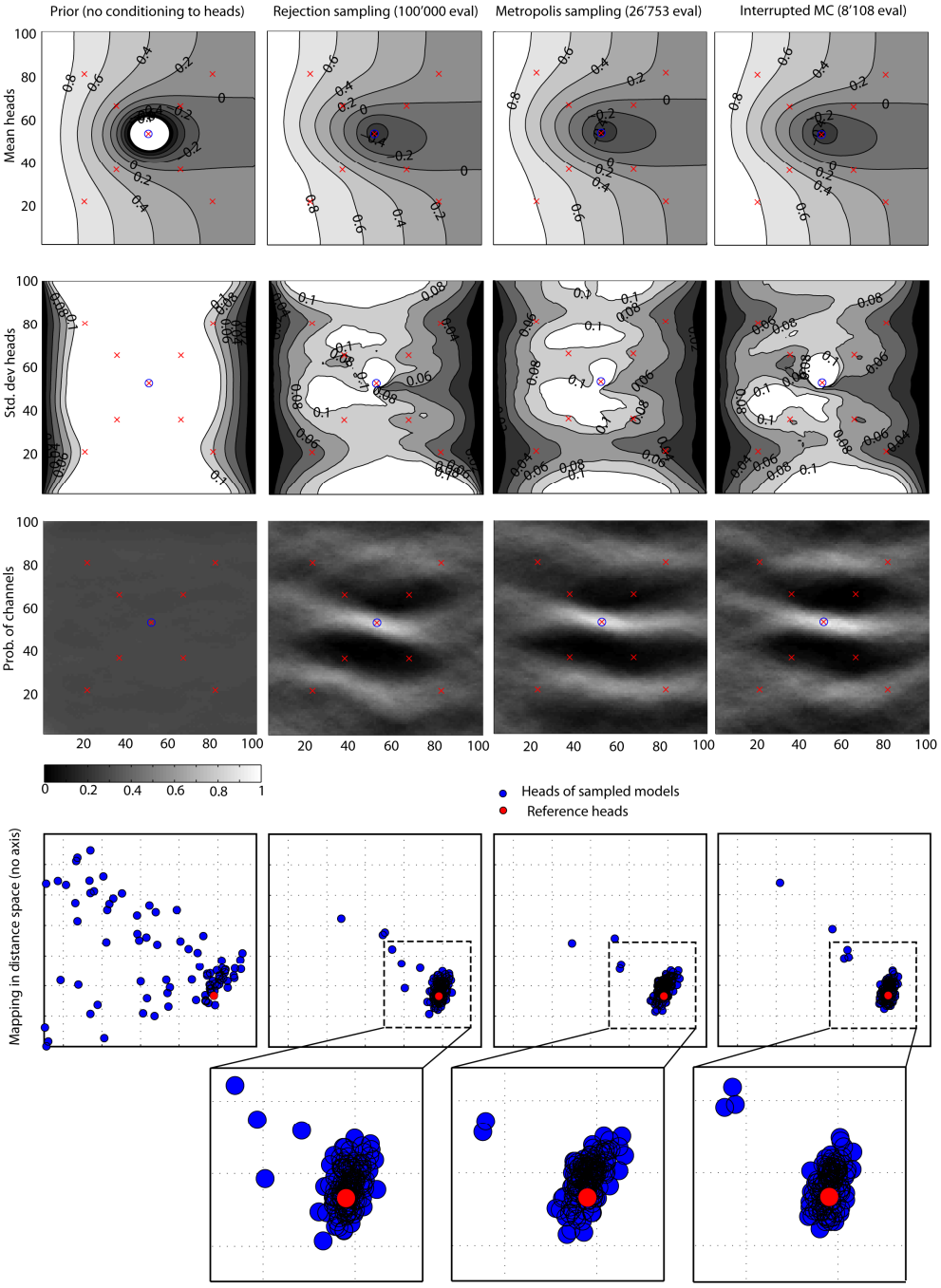

Figure 7. Representation of the ensembles of models obtained with different methods (150 models per method). Each column represents the results of a sampling method. Each row corresponds to a different representation of the ensemble.

The first column of Figure 7 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 the head is not represented below -1 m). Since no conditioning hard 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.

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

$$L(\mathbf{m}_i) = \exp\left(-\frac{\mathrm{RMSE}(\mathbf{m}_i)^2}{2\sigma^2}\right), \tag{5}$$

with σ = 0.03 m, which can reasonably correspond to the measurement error on head. The supremum value corresponds to a RMSE of 0.0300. After 100'000 evaluations, 150 realizations are sampled, representative of $f(\mathbf{m}|\mathbf{d})$. Nine of these realizations, with their respective fits to data, are displayed in Figure 8. Although showing good fits, they are very different. This is an indication of non-uniqueness of the solutions and of multiple minima in the solution space.

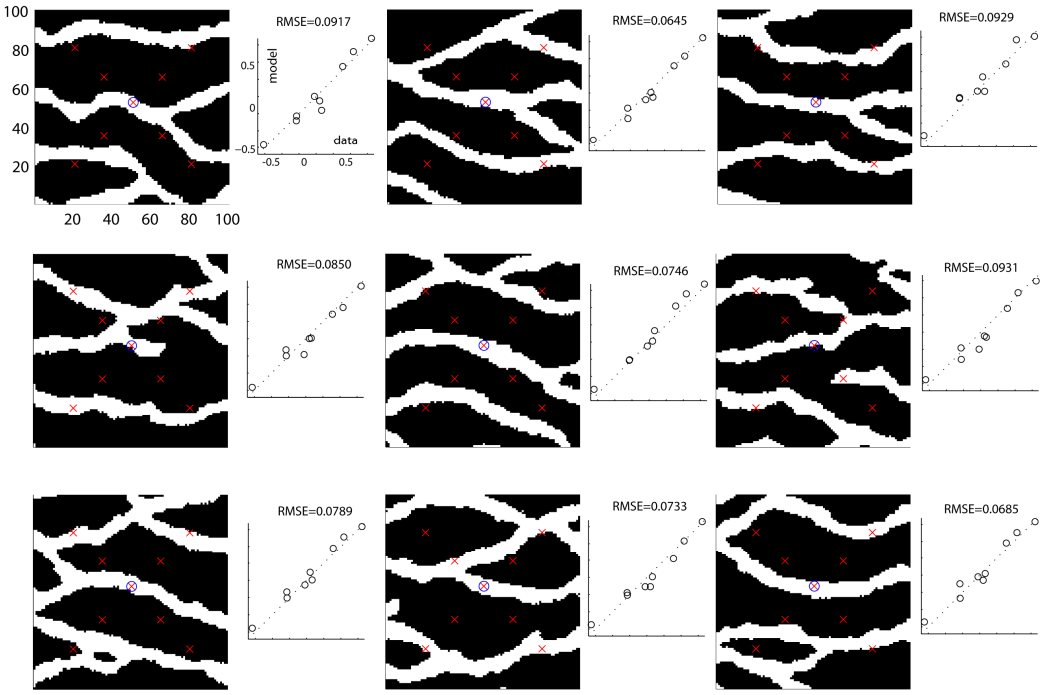

Figure 8. Nine realizations out of the 150 sampled with the rejection method. The

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 are only displayed for the top left image.

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 downwards 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.

Now that the posterior distribution is entirely characterized with rejection, we perform another sampling using a Metropolis sampler as described in section 2.2, with likelihood (5). A constant resampling factor of φ=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 9. It shows the mean RMSE of the stack of models accepted by the Metropolis algorithm as a function of the stack size. Upon convergence, variations between the RMSE of samples are expected to cancel out and the mean RMSE should stabilize. Using Figure 9, 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.

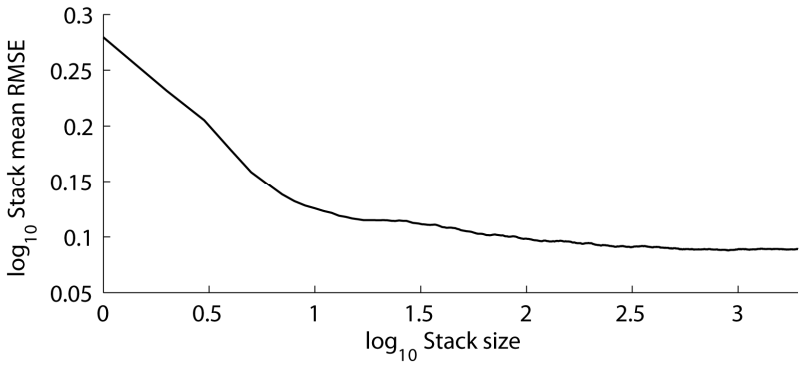

Figure 9. Convergence of the Metropolis sampler in the synthetic test case.

The third column of Figure 7 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 7 outliers and Metropolis only 4. Moreover,

Metropolis sampling results in 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.

The fourth column of Figure 7 represents 150 samples obtained by interrupted Markov chains, with a constant fraction of resampled nodes of φ=0.01. Likelihood (5) 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 8'108 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, 5 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 part). Figure 10 shows the evolution of the 150 optimizations and their interruptions. The number of iterations i before interruption range between 4 and 144, with an average of 54 iterations. Nine optimized realizations obtained by interrupted Markov chains are shown in Figure 11. Similarly to the case of rejection sampling, the presence of diversity in the population of solutions indicates that different local minima have been explored.

For comparison, Figure 12 displays the distance-based representation of models obtained 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.

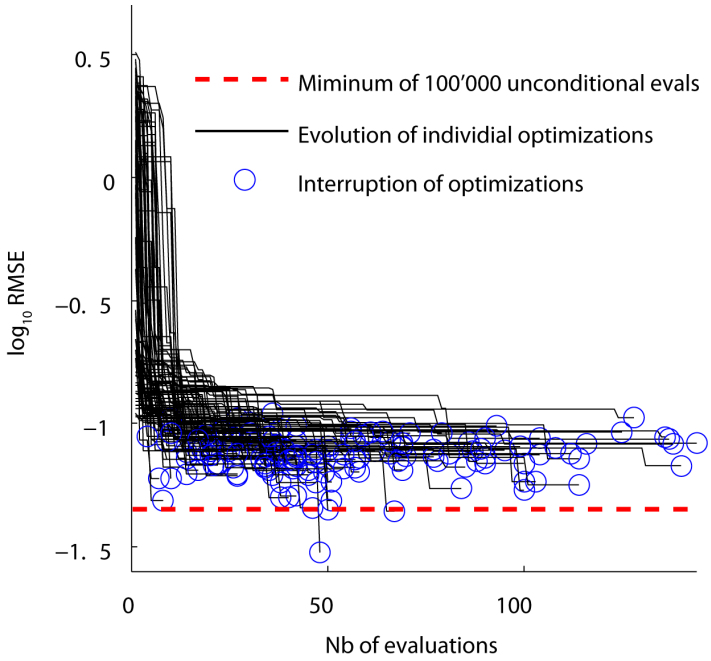

Figure 10. Evolution of the 150 individual optimizations used for interrupted

Markov chains.

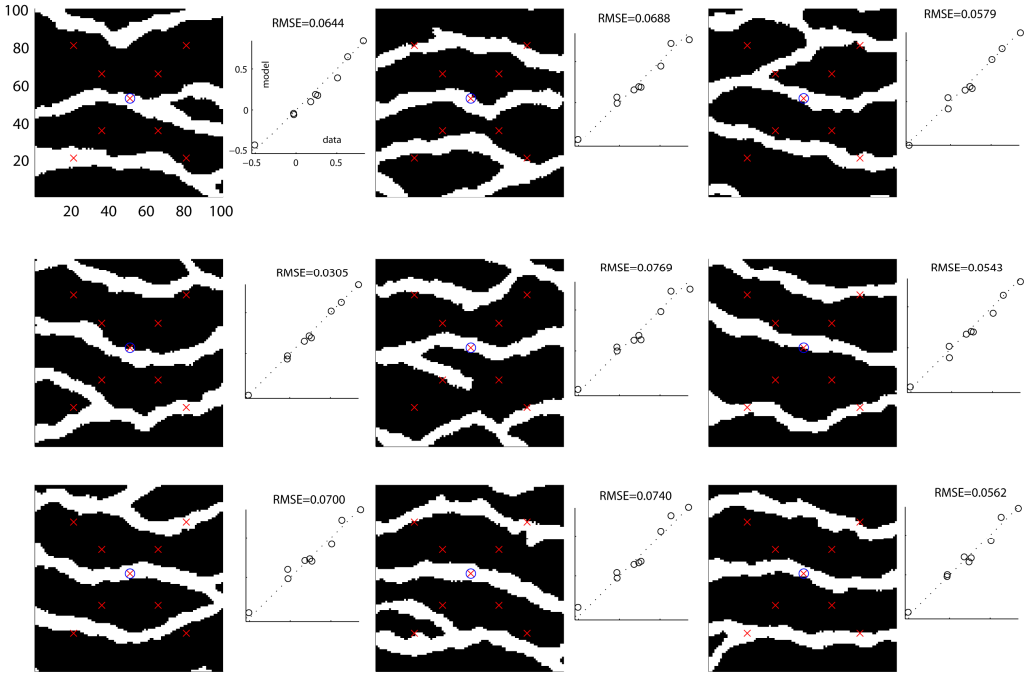

Figure 11. Nine realizations out of the 150 sampled with interrupted Markov chains. The fits to data and the RMSE are shown on the right of each realization. The axes and labels are the same for all realizations, but are only displayed for the top left image.

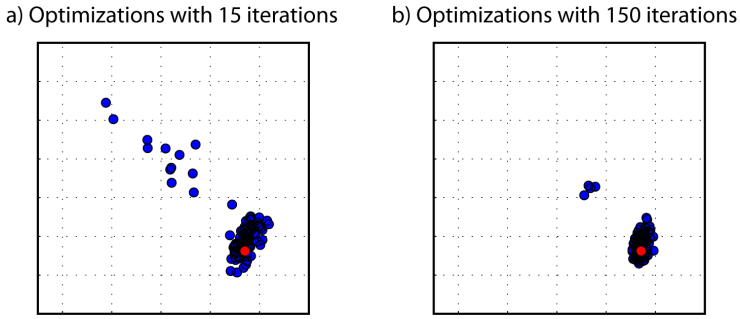

Figure 12. Distance-based representation of ensembles obtained with deterministic stopping criteria. a) Fixed number of 15 iterations. b) Fixed number of 150 iterations.

4. Conclusions

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 acceptation/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 the time. ISR can therefore be applied in both contexts of Bayesian inversion and optimization. For this 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 samplers, with important reduction in CPU cost.

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 implemented without modification of existing computer codes. The method is simple in its concept and needs very little parameterization.

The fraction of resampled nodes φ is the only parameter required by ISR. It has been shown that optimization with ISR is efficient for a wide range of φ. 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 the optimal parameter.

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. 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.

5. Bibliography

Alcolea, A., et al. (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Advances in Water Resources*, *2006*(29), 1678-1689.

Alcolea, A., and P. Renard (in press), The Blocking Moving Window sampler. Conditioning Multiple-Point Simulations to Hydrogeological data., *Water Resour. Res.*

Ballin, P., et al. (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, Louisiana.

Bayer, P., et al. (2007), Computationally efficient stochastic optimization using multiple realizations, *Advances in Water Resources*, *31*(2), 399-417.

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

Boucher, A. (2007), Algorithm-driven and representation-driven random function : A new formalism for applied geostatistics, edited, Stanford Center for Reservoir Forecasting, Palo Alto, CA.

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

Caers, J. (2005), *Petroleum Geostatistics*, 88 pp., Society of Petroleum Engineers, Richardson.

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

Caers, J. (2007), Comparing the Gradual Deformation with the Probability Perturbation Method for Solving Inverse Problems, *Mathematical Geology*, *39*(1), 27-52.

Carrera, J., et al. (2005), Inverse problem in hydrogeology, *Hydrogeology Journal*, *13*, 206-222.

De Marsily, G., et al. (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. (ed), pp. 831-849, D. Reidel Pub.

De Marsily, G., et al. (2005), Dealing with spatial heterogenity, *Hydrogeology Journal*, *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, *Stoch Envir Res and Risk Ass*, *14*(2004), 401-413.

Fraser, A. (1957), Simulation of Genetic Systems by Automatic Digital Computers. I. Introduction, *Australian Journal of Biological Sciences*, *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 Problems in Science and Engineering*, *16*(7), 865-884.

Fu, J., and J. Gomez-Hernandez (2009a), A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling, *Math. Geosci.*, *2009*(41), 105-128.

Fu, J., and J. Gomez-Hernandez (2009b), Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method, *Journal of Hydrology*, *2009*(364), 328-341.
Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, 412 pp., Addison-Wesley, Berlin.
Gomez-Hernandez, J., et al. (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. Theory, *Journal of Hydrology*(203), 162-174.
Guardiano, F., and M. Srivastava (1993), Multivariate geostatistics: Beyond bivariate moments, in *Geostatistics-Troia*, edited by A. Soares, pp. 133-144, Kluwier Academic, Dordrecht.
Hendricks-Franssen, H.-J., et al. (2004), Joint estimation of transmissivities and recharges—application: stochastic characterization of well capture zones, *Journal of hydrology*, *294*(2004), 87-102.
Hernandez, A., et al. (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, *Mathematical Geology*, *32*(1), 87-108.
Hu, L., et al. (2001), Gradual Deformation and Iterative Calibration of Sequential Stochastic Simulations, *Mathematical Geology*, *33*(4), 475-489.
Johansen, K., et al. (2007), Hybridization of the probability perturbation method with gradient information, *Computers & Geosciences*, *2007*(11), 319-331.
Journel, A., and T. Zhang (2006), The Necessity of a Multiple-Point Prior Model, *Mathematical Geology*, *38*(5), 591-610.
Karpouzos, D., et al. (2001), A Multipopulation Genetic Algorithm to Solve the Inverse Problem in Hydrogeology, *Water Resour. Res.*, *37*(9), 2291-2302.
Kirkpatrick, S., et al. (1983), Optimization by Simulated Annealing *Science*, *220*(4598), 671-680.
Kjønsberg, H., and O. Kolbjørnsen (2008), Markov mesh simulations with data conditioning through indicator kriging, paper presented at Geostats 2008, 1-5 Dec. 2008.
Le Ravalec-Dupin, M., and B. Noetinger (2002), Optimization With the Gradual Deformation Method, *Mathematical Geology*, *34*(2), 125-142.
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 Science and Technology*, *62*(2), 169-180.
Liu, N., and D. Oliver (2004), Experimental Assessment of Gradual Deformation Method, *Mathematical Geology*, *36*(1), 65-77.
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), *Journal of hydrology*, *371*(2009), 75-84.
Llopis-Albert, C., and J. Capilla (2009), Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 2. Demonstration on a synthetic aquifer, *Journal of hydrology*, *371*(2009), 53-65.

Mariethoz, G., and P. Renard (2010), Reconstruction of incomplete data sets or images using Direct Sampling, *Mathematical Geosciences*, *42*(3), 245-268.
Mariethoz, G. (in press), A general parallelization strategy for random path based geostatistical simulation methods, *Computers & Geosciences*.
Mariethoz, G., et al. (submitted), The direct sampling method to perform multiple-points simulations, *Water Resour. Res.*
Metropolis, N., et al. (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, *Journal of geophysical Research*, *100*(B7), 12,431-412,447.
Oliver, D., et al. (1997), Markov chain Monte Carlo methods for conditioning a logpermeability field to pressure data, *Math. Geosci.*, *29*(1), 61-91.
Omre, H., and H. Tjelmland (1996), Petroleum geostatistics, Department of Mathematical Sciences, Norwegian University of Science and Technology, 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., et al. (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., et al. (2009), *Applied Geostatistics with SGeMS: A User's Guide*, 284 pp., Cambridge University Press, Cambridge.
Ronayne, M., et al. (2008), Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach, *Water Resour. Res.*, *44*(W08426).
Scheidt, C., and J. Caers (2009), Representing Spatial Uncertainty Using Distances and Kernels, *Mathematical Geosciences*, *41*(2009), 397-419.
Shannon, C. E. (1948), A mathematical theory of communication, *The Bell system technical journal*(27), 379-423.
Singh, A., et al. (2008), An interactive multi-objective optimization framework for groundwater inverse modeling, *Advances in Water Resources*, *31*(2008), 1269-1283.
Smith, J. (1997), Quick Simulation: A Review of Importance Sampling Techniques in Communications Systems, *IEEE journal on selected areas in communications*, *15*(4), 597-613.
Strebelle, S. (2002), Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, *Mathematical Geology*, *34*(1), 1-22.
Subbey, S., et al. (2004), Prediction under uncertainty in reservoir modeling, *Journal of Petroleum Science and Engineering*, *44*(2004), 143-153.
Tarantola, A. (2005), *Inverse Problem Theory and Methods for Parameter estimation*, Society for Industrial and Applied Mathematics, Philadelphia.
Tsai, F., et al. (2003), Global-local optimization for parameter structure identification in three-dimensional groundwater modeling, *Water Resour. Res.*, *39*(2), 1043-1057.

Vesselinov, V., et al. (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, *Nat. Bureau Standards*, *12*, 36–38.
Yeh, W. (1986), Review of Parameter Identification Procedures in Groundwater Hydrology : The Inverse Problem, *Water Resour. Res.*, *22*(2), 95-108.
Zheng, C., and P. Wang (1996), Parameter structure identification using tabu search and simulated annealing, *Advances in Water Resources*, *19*(4), 215-224.