Bayesian updating of models parameters: a first step towards solving inverse problems

Gregoire Mariethoz, Jef Caers

Department of Energy Resources Engineering
Stanford University

Abstract

Modern Earth Science models are complex and highly parameterized. While offering wide possibilities for modeling realistic geological features, this complexity can involve difficulties in the inference of parameters. One often knows ranges of possible values for individual parameters, but jointly adjusting them can be challenging and time-consuming.

In this context, we propose a general methodology to identify what combinations of parameters are best suited to match certain data. It relies on a Bayesian framework and uses distance-based techniques. We present the approach as a general way of organizing the parameters inference effort, which could greatly enhance the efficiency and the accuracy of inverse modeling techniques.

The approach is illustrated with a simple object-based model. The methodology allows obtaining combinations of parameters that are more likely to match the data, even though parameters taken individually are uncorrelated to the data.

1. Introduction

When building models to characterize the subsurface, one can distinguish two main sources of uncertainty: *spatial uncertainty* and *parameters uncertainty*. In the frame of inverse problems, emphasis is often put on spatial uncertainty. Given certain statistical parameters of geological continuity, one wishes to find realizations of a variable that correspond to available data (hard, soft, dynamic data). However, significant uncertainty may be associated to the parameters defining the statistical model of geological continuity. Ignoring this uncertainty can result in inadequate model parameters, and hence inaccurate and inefficient resolution of inverse problems. Actually, considering only spatial uncertainty tends to give a false sense of

comfort that the inverse solutions are correct since they are matching the data. Solving inverse problems is more than just producing models that match the data. Solutions are also required to reproduce some prior statistics and to be sampled correctly from the posterior as imposed by Bayes' rule [*Mosegaard and Tarantola*, 1995; *Omre and Tjelmland*, 1996; *Tarantola*, 2005]. Therefore, prior to using any stochastic inverse modeling techniques aiming at resolving spatial uncertainty [e.g. *Caers and Hoffman*, 2006; *Doherty*, 2003; *Fu and Gomez-Hernandez*, 2008; *Hu*, 2000; *Kitanidis*, 1995; *Mariethoz et al.*, submitted], it is important to appropriately define ranges of model parameters in a Bayesian framework.

Although our approach is general, we will focus on the parameters used by geostatistical methods to define spatial variation. The resulting parameters can be related to a variogram model (mean, variance, variogram type and range), to the objects characteristics and rules used in a Boolean model [*Lantuéjoul*, 2002] or to the choice of a training image (scenario A, B or C) for multiple-point methods [*Guardiano and Srivastava*, 1993]. In all cases, a combination of several parameters is required. Most often, interaction between parameters is expected. For example in the Boolean case, the proportion of a certain facies is dependent of both the size and the number of objects representing this facies. Investigating all possible combinations of parameters is often not feasible since it requires building and evaluating a prohibitive amount of models. In simple cases, it is possible to empirically identify combinations of parameters that can be immediately discarded. In the Boolean example if a large proportion a facies is desired, it is obvious that combinations involving a small number of small objects should be avoided.

However, with modern Earth Sciences models of increasing complexity, empirically guessing the appropriate combinations of parameters is not feasible. For example, pseudo-genetic geological models [*Deutsch and Tran*, 2002; *Michael et al.*, in press; *Zhang et al.*, 2009] need parameters defining the shapes of objects, their size and density, and possibly rules determining the stacking or erosion of objects. In such cases, the number of parameters is large and the multiple interactions are intractable. Moreover, with the increasing prominence of algorithm-driven random functions [*Boucher*, 2007; *Journel and Zhang*, 2006], the effect of certain parameters on the final model are not easily grasped from a qualitative point of view.

In inverse modeling, one attempts to create 3D models that match the data by varying both input parameters and spatial uncertainty. Since most inverse problems are ill-posed, several such models can be created. The posterior variation of the parameters corresponding to the inverse solutions constitutes the posterior parameters distribution. Two difficulties arise in this attaining this posterior

1) Defining a prior distribution for the input parameters (a range of prior input parameters) that is realistic and compatible with the data.
2) Solving the inverse problem for varying parameters as well as accounting for spatial uncertainty. This is difficult from a methodological point of

view as well as computationally.

The contribution of this paper is to propose a formal and general probabilistic approach for determining the posterior input parameters without doing any explicit inverse modeling. The approach relies on distance-based techniques [*Scheidt and Caers*, 2009a; b; 2010; *Suzuki and Caers*, 2008] that define a distance (a single scalar value) that characterizes the dissimilarity between any two objects (for example geographical locations, geological models or sets of parameters). The purpose of this distance is to bring structure to the large variability of geological models.

We introduce a parameters distance that defines the dissimilarity between combinations of parameters used to generate models. Although the inference of models parameters is often not posed in terms of distance, any adjustment of parameters, using any currently available technique, requires the definition of a metric between parameters or parameters sets. For example, when fitting the range of a variogram on field-scale measurements, one knows that proceeding by millimetric adjustments is not adequate and that it is more efficient to test values that are further apart. A notion of distance is also implicitly present when dealing with permeability that is usually log-transformed in order to carry on data analysis and modeling. Such a transformation is nothing else than a redefinition of the distances between permeability values. The new distance produces a parameters space easier to deal with. Any subsequent comparison between permeability values is only valid in this new distance space. These examples show that the inference of parameters is only possible if a distance between parameters is (implicitly or explicitly) defined.

There is a direct equivalence between the probability density function (pdf) of parameters and the density of these parameters in a metric space. Hence it is straightforward to obtain high-dimensional prior distributions for combinations of parameters. More precisely, we use low-dimensional projections of metric spaces using multi-dimensional scaling (MDS). Then, Bayes' theorem can be applied to update the prior distribution of parameters into a posterior distribution that is constrained by the data [see *Tarantola*, 2005 for a review on Bayesian methods]. Drawing parameters sets from this posterior distribution allows generating new models that provide increased match to data. Since the posterior distribution is focused towards reproduction of the data, it represents a much smaller uncertainty space to search in the frame of inverse problems.

2. Methodology

The overall methodology consists in generating, in a first step, a series of models using prior combinations of parameters. Once these prior models have been evaluated, Bayesian updating allows narrowing the combinations of parameters that are compatible with the data. Note that no inverse or optimization problem is solved explicitly. In a second step, new models are generated generate using the updated combinations of parameters. Central to the methodology is the use of distance-based

techniques to infer high-dimensional probability distributions.

2.1. Computing a distance between parameters

Using Multi-Dimensional Scaling or MDS [*Borg and Groenen*, 1997] with an appropriate metric, one can create a map of a set of combinations of parameters. Given a distance matrix **D**, such a representation displays an ensemble of parameters combinations $\boldsymbol{\theta}_i$ as a set of points $\mathbf{x}_i$ in a possibly high-dimensional Euclidean space $\Re$, each point corresponding to a combination of parameters. Points in $\Re$ are 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. Our approach relies on the definition of a specific distance that is tailored to represent the dissimilarities between combinations of parameters.

The parameter vectors $\boldsymbol{\theta}$ used for geological models can be of a heterogeneous nature, in the sense that some parameters can be categorical and other can be continuous. Categorical parameters do not have an order relationship. For example, they relate to the choice of a training image in the context of multiple-point simulations or the type of objects for Boolean methods. Continuous parameters, on the other hand, can also have a large impact. This is the case for the size and the number of objects, or for the distance threshold in the direct sampling multiple-point simulation method [*Mariethoz and Renard*, 2010].

Let us first consider the distance between individual parameters θ. If the parameter considered is continuous, we compute the distance between any two parameters values θ^1 and θ^2 as

$$d\{\theta^1,\theta^2\}=0 \text{ if } \theta^1=\theta^2, \text{otherwise } d\{\theta^1,\theta^2\}=1. \quad (1)$$

If the parameter is continuous or if it has an order relationship, the distance is defined as

$$d\{\theta^1,\theta^2\}=\eta\left|\theta^1-\theta^2\right|, \quad (2)$$

were η is a normalization factor chosen such that $d\{\theta^1,\theta^2\}$ is comprised in the interval [0 1]. Note that these distances represent one specific choices among other possibilities. For example, one could choose an Euclidean distance for continuous variables.

To compute a distance between combinations of parameters, let us consider vectors of parameters $\boldsymbol{\theta}^1$ and $\boldsymbol{\theta}^2$, where $\boldsymbol{\theta}^1=\{\theta_1^1,...,\theta_n^1\}$ and $\boldsymbol{\theta}^2=\{\theta_1^2,...,\theta_n^2\}$. The distance between parameters vectors $d\{\boldsymbol{\theta}^1,\boldsymbol{\theta}^2\}$ can be defined as a linear combination of the distances between each individual parameter in the parameters vectors:

$$d\{\boldsymbol{\theta}^1,\boldsymbol{\theta}^2\}=\frac{1}{n}\sum_{j=1}^{n} d\{\theta_j^1,\theta_j^2\}. \quad (3)$$

Since a distance is always a continuous quantity, it organizes combinations of categorical and continuous parameters. This allows formalizing concepts of similarity and dissimilarity between heterogeneous parameters vectors and making the problem more structured and continuous, an important characteristic of the distance-based methods.

2.2. Bayesian updating of prior parameters

Once a distance between parameters is defined, one can generate P parameters vectors $\boldsymbol{\theta}_p$, $p=\{1,\ldots,P\}$ from any given prior distribution (independent univariate, multivariate, etc) and map them in $\Re$ using MDS. Since no data have been introduced at this stage, the resulting locations $\mathbf{x}_p$ after projection represent an empirical representation of the prior parameters distribution $f(\mathbf{x}_p)$, provided that P is large enough to adequately cover the prior. More precisely, the density of points $\mathbf{x}_p$ will be higher where parameters combinations often occur and lower where they seldom appear.

Next, we attribute a likelihood $L(\boldsymbol{\theta}_p)$ to each combination of parameters. This requires creating a geological model with the given parameters combination and running the forward model. Note we create a single geological model for each combination of parameters. In the next section, we show that since a distance between parameters is defined, it is still possible to evaluate the variability of models for a given parameters combination.

The likelihood represents the misfit between the observed data and the prediction of a model. Computing the misfit e_p requires running a forward model of the form

$$e_p=g\{m(\boldsymbol{\theta}_p)\}, \quad (4)$$

which means that parameters cannot be evaluated directly, but are used to generate a model $m(\boldsymbol{\theta}_p)$, which will itself be used to compute the error relative to the data. The likelihood is then computed as a function of e_p. Since there is a one-to-one correspondence between $\boldsymbol{\theta}_p$ and $\mathbf{x}_p$, $L(\boldsymbol{\theta}_p)$ can be directly associated to $L(\mathbf{x}_p)$.

With the prior distribution and the likelihood that are defined, Bayes' rule can be used to define the posterior probability of each combination of parameters as

$$f(\mathbf{x}_p \mid \mathbf{d}) \sim f(\mathbf{d} \mid \mathbf{x}_p) f(\mathbf{x}_p), \quad (5)$$

where $\mathbf{d}$ represents the data constraining the inverse problem and $f(\mathbf{d}|\mathbf{x}_p)$ denotes the likelihood of a given combination of parameters. One can therefore evaluate the posterior probability of all prior points $\mathbf{x}_p$.

Using density interpolation methods such as kernel smoothing [*Epanechnikov*,

1969], it is possible to estimate $f(\mathbf{x}_p)$ at any location $\mathbf{y}$ in $\Re$ based on the density of prior samples $\mathbf{x}_p$. The density at $\mathbf{y}$ is computed by summing kernel functions that are centered on each prior sample. If a Gaussian kernel of standard deviation σ_k is used, the value at location $\mathbf{y}$ of a kernel centered on $\mathbf{x}_p$ is

$$w_p(\mathbf{y}) = \frac{\exp\left(\frac{-\|\mathbf{y}-\mathbf{x}_p\|^2}{2\sigma_k^{\,2}}\right)}{\sum_{i=1}^{P}\exp\left(\frac{-\|\mathbf{y}-\mathbf{x}_i\|^2}{2\sigma_k^{\,2}}\right)}. \tag{7}$$

The prior distribution at any location $\mathbf{y}$ is given by

$$f(\mathbf{y}) = \sum_{p=1}^{P} w_p(\mathbf{y}), \tag{6}$$

and the posterior density at $\mathbf{y}$ estimated using Bayes' rule is

$$f(\mathbf{y} \mid \mathbf{d}) \sim \sum_{p=1}^{P}\left\{w_p(\mathbf{y})L(\mathbf{x}_p)\right\}, \tag{8}$$

which can be seen as a weighted mean of the likelihood where the weights are given by the Gaussian kernel.

2.3. Distinguishing spatial uncertainty and parameters uncertainty

In most cases, a combination of parameters $\boldsymbol{\theta}_p$ can yield several possible errors values e_p. This non-uniqueness of the forward model (eq. 4) is due to the addition of a random component in the computation of $m(\boldsymbol{\theta}_p)$. For example, sequential simulation methods randomize the simulation path and Boolean methods randomize the location of individual objects.

If this random component plays a significant role, it can result in a large spatial uncertainty. Characterizing spatial uncertainty could be accomplished by generating several realizations with each prior parameters vector $\boldsymbol{\theta}_p$ and performing a variance analysis on the resulting models. The drawback of this approach is that it would only provide information on spatial uncertainty at locations where prior parameters are sampled (i.e. not at an arbitrary unsampled location $\mathbf{y}$). Moreover, it would require a large number of forward problem runs.

Eq. 8 allows for the estimation of the likelihood at $\mathbf{y}$ by a weighted mean, but does not provide a measure of the likelihood variation in the neighborhood of $\mathbf{y}$. We propose to quantify this likelihood variation by a weighted variance

$$v(\mathbf{y}) = \sum_{p=1}^{P} w_p(\mathbf{y})\left\{L(\mathbf{x}_p) - L(\mathbf{y})\right\}^2 . \quad (9)$$

that has the same weights w_p as those used for the weighted mean (eq. 8). $v(\mathbf{y})$ is a measure of the likelihood variability for similar combinations of parameters, and therefore gives a direct indication of the spatial uncertainty. In areas where $v(\mathbf{y})$ is large, spatial uncertainty dominates, whereas parameters uncertainty is more significant where $v(\mathbf{y})$ is low.

2.4. Sampling the posterior parameters

Now that a posterior distribution of parameters is accessible via eq. 8, it is possible to generate new combinations of parameters that are distributed according to the estimated $f(\mathbf{x}_p|\mathbf{d})$. To distinguish them from the prior samples $\mathbf{x}$, these new posterior samples are denoted $\mathbf{y}$. Each sample $\mathbf{y}$ is the image in $\Re$ of a combination of posterior parameters that is more likely than the prior parameters combinations to produce models honoring the data.

According to Mosegaard and Tarantola [1995], the Metropolis sampler is an adequate way of sampling distributions whose functional form is unknown. To apply it, one needs to design a random walk that samples the prior by proposing, at each step i, a new location $\mathbf{y}^*$. The random walk moves according to the following rules:

- if $L(\mathbf{y}^*) \geq L(\mathbf{y}_i)$, move from $\mathbf{y}_i$ to $\mathbf{y}^*$.
- if $L(\mathbf{y}^*) < L(\mathbf{y}_i)$, randomly choose to move to $\mathbf{y}^*$ or stay at $\mathbf{y}_i$, with the probability $L(\mathbf{y}^*) / L(\mathbf{y}_i)$ of moving to $\mathbf{y}^*$.

The movement (or transition) from a $\mathbf{y}_i$ to $\mathbf{y}^*$ is accomplished by drawing a candidate location $\mathbf{y}^*$ from the proposal distribution $Q(\mathbf{y}^*|\mathbf{y}_i)$, which denotes the probability density function of the transition from the $\mathbf{y}_i$ to $\mathbf{y}^*$. The method requires the proposal density to be symmetric, i.e. that $Q(\mathbf{y}_i|\mathbf{y}^*) = Q(\mathbf{y}^*|\mathbf{y}_i)$. Adding a to $\mathbf{y}_i$ a random uniform noise $\mathbf{u}$ is a movement that honors this condition:

$$\mathbf{y}^* = \mathbf{y}_i + \mathbf{u} . \quad (10)$$

When such a random walk has reached its steady-state, the accepted locations are guaranteed to be samples of $f(\mathbf{x}_p|\mathbf{d})$. The Metropolis sampler can be used to obtain samples in high-dimension spaces, however, to increase its efficiency we ignore the dimensions carrying less information. Note that for the Metropolis sampler to be efficient, the magnitude of $\mathbf{u}$ should be small compared to the magnitude of $\mathbf{y}_i$, which means relatively small steps for the random walker.

2.5. Pre-image problem

The Metropolis sampler allows quickly generating Q samples $\mathbf{y}_q$, $q=\{1,\ldots,Q\}$.

Finding the actual posterior combinations of parameters consists in finding $\boldsymbol{\theta}_q$ such that its image in $\Re$ is $\mathbf{y}_q$. This problem is often referred to as a pre-image problem [*Mika et al.*, 1999], and can be solved using optimization techniques. Note that the solution of the pre-image problem can be non-unique.

Since $\Re$ is defined by a distance between a set of prior points $\mathbf{x}_p$, the location of a point $\mathbf{y}_q$ in $\Re$ can be entirely characterized by the distance between $\mathbf{y}_q$ and all prior points $\mathbf{x}_1,\ldots,\mathbf{x}_p$. For a given image $\mathbf{y}_q$, we do not directly know $\boldsymbol{\theta}_q$, but we know the distance between $\mathbf{y}_q$ and all prior models $\mathbf{x}_p$. Therefore solving the pre-image problem amounts to finding $\boldsymbol{\theta}_q$ such that $d(\boldsymbol{\theta}_q,\boldsymbol{\theta}_p)$ is proportional to $d(\mathbf{y}_q,\mathbf{x}_p)$ for all p. The objective function to minimize is then

$$o_q = \sum_{p=1}^{P}\left(1-\frac{d(\boldsymbol{\theta}_q,\boldsymbol{\theta}_p)}{d(\mathbf{y}_q,\mathbf{x}_p)}\right). \tag{11}$$

This can be accomplished by a number of methods. In this paper, we use the Nelder-Mead simplex method [*Lagarias et al.*, 1998; *Nelder and Mead*, 1965]. In addition, the optimization procedure is designed to ensure that the posterior parameters vectors remain within the bounds of the prior parameters. As a result, we obtain parameters combinations $\boldsymbol{\theta}_q$ that are more targeted than the prior parameters $\boldsymbol{\theta}_p$. Using these new, updated combinations of parameters should yield models of increased likelihood.

3. Example

3.1. Problem setting

To validate the proposed methodology, we design a synthetic test case consisting of a binary variable Z discretized on a square domain of 81 by 81 pixels. Realizations of Z are produced using a variety of parameters. We first define a prior distribution for these parameters, and then we update it using synthetic data. Finally, we generate new realizations with the updated parameters distribution.

Realizations are obtained using a mixture of different methods, with parameters that are interpreted differently for each method. They are generated with a stochastic simulator that has five possible parameters explained below:

α : the type of model of spatial continuity. 4 possible types are considered: 1) Boolean model using ellipses objects, 2) Boolean model using lines objects, 3) truncated Gaussian simulation, 4) Boolean model using rectangle objects.

β : anisotropy of the structures along the X axis.

γ : anisotropy of the structures along the Y axis.

δ : size of objects.

ε : number of objects.

All object-based models types (α = 1,2,4) consist in randomly populating the grid with ε identical objects defined by the parameters β, γ and δ. We assign Z=1 at locations within objects and Z=0 elsewhere. For ellipses (α = 1) and rectangle objects (α = 4), the anisotropy ratio of objects is the ratio β/γ, and the area of the objects is proportional to δ. For lines objects (α = 2), the slope dx/dy is defined as 2(γ - 0.5), and the parameter β is ignored. Since in our example γ varies in the interval [0 1], the slope of the lines ranges between -2 and 2.

Truncated Gaussian simulations (α = 3) are based on a multiGaussian field Z^* of zero mean and unit variance, generated using an exponential variogram model with correlation ranges $\beta \cdot \delta$ in the X direction and $\gamma \cdot \delta$ in the Y direction. The parameter ε is then ignored. Truncation is performed by assigning Z=1 at locations where Z^*>0.5 and Z=0 elsewhere.

For a given set of input parameters $\boldsymbol{\theta}$=[α, β, γ, δ, ε], the stochastic simulator produces a field of objects. This test case is designed to study interactions between estimated parameters. For example, parameter δ (size of objects) generally influences the proportion of facies 1, but this influence will largely depend on α (except if α = 3, then δ is ignored), and on ε (number of objects). For example a low ε value can balance a high δ value. Figure 1a shows one realization with different parameters sets, and Figure 1b displays the parameters combinations used for each of these 4 realizations.

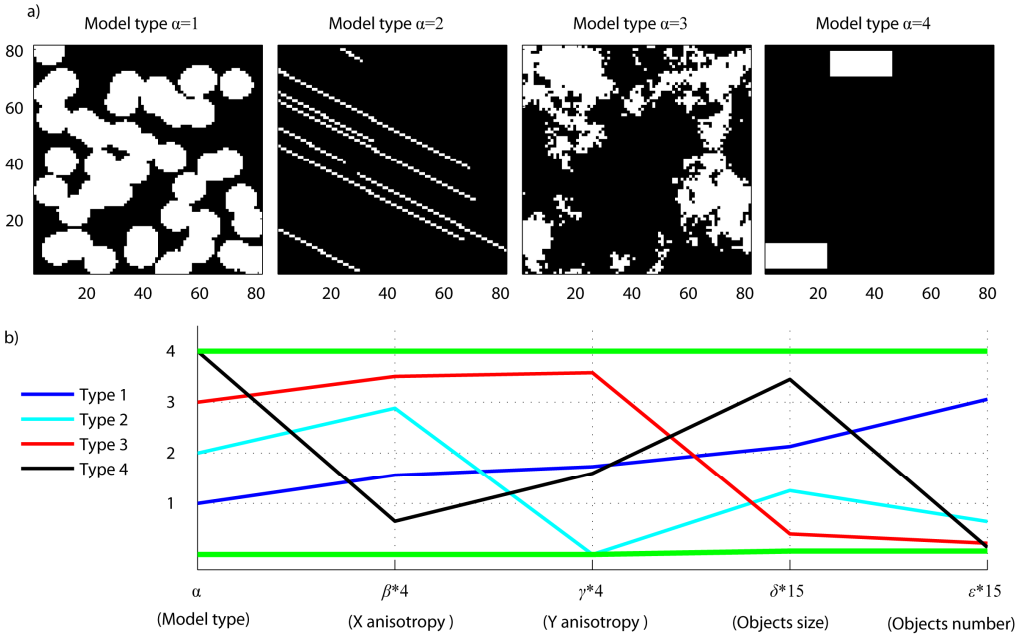

Figure 1. Realizations using different combinations of parameters. a) One realization with each model type. b) representation of the parameters combinations

used to generate the realization shown in a). Each combination is displayed as a line of a different color. The abscissa represents the parameters and the ordinate represents the scaled value taken by each parameter. Parameters values are scaled for representation. The green line represents the bound on the range of parameter values

We define a reference set of parameters, $\boldsymbol{\theta}_{ref}$ = [2, 0.6, 0.4, 30, 45] as our so-called "truth". One realization of Z is generated using $\boldsymbol{\theta}_{ref}$ and considered as the reference field $\mathbf{m}_{ref}$. The reference is used to produce synthetic data, but is otherwise considered unknown in the rest of the paper. Data $\mathbf{d}_{ref}$ are sampled from this reference, and consist in the ensemble of pixels located in a small zone (21 by 21 pixels) in the center of the reference realization, over which a moving average has been applied. The moving average is computed at each data pixel by averaging the values of the neighbors of in a window of 9 by 9 pixels. In a real setting, $\mathbf{d}_{ref}$ would correspond to the probability of occurrence of facies 1, an information derived from geophysical surveys. Figure 2a displays the reference field $\mathbf{m}_{ref}$ with the data zone represented as a red square. The reference data $\mathbf{d}_{ref}$ is shown in the zoomed area. All models subsequently produced will be evaluated by computing their mismatch with respect to this local moving average.

We generate P=100 prior parameters sets $\boldsymbol{\theta}_p$ by drawing them from $f(\boldsymbol{\theta})$, defined as uniform distributions that are independent for each parameter and whose bounds are given in Table 1.

	$f(\alpha)$	$f(\beta)$	$f(\gamma)$	$f(\delta)$	$f(\varepsilon)$
parameter description	object type	X anisotropy	Y anisotropy	objects size	nb. of objects
parameter type	categorical	continuous	continuous	continuous	continuous
upper bound	4	1	1	60	60
lower bound	1	0	0	1	1
normalization factor η	n/a	1	1	1/59	1/59

Table 1. Significance, type, normalization and prior distribution for each of the five parameters used to generate realizations of Z.

A forward problem (eq. 4) is computed for each vector of parameters $\boldsymbol{\theta}_p$,

consisting in 1) generating a model $m(\boldsymbol{\theta}_p)$ and 2) computing the error e_p as the RMSE between $\mathbf{d}_p$ and $\mathbf{d}_{ref}$, with $\mathbf{d}_p$ defined as the moving average of $m(\boldsymbol{\theta}_p)$ in the data zone. Figure 2b and Figure 2c show the models presenting respectively the smallest and the largest error.

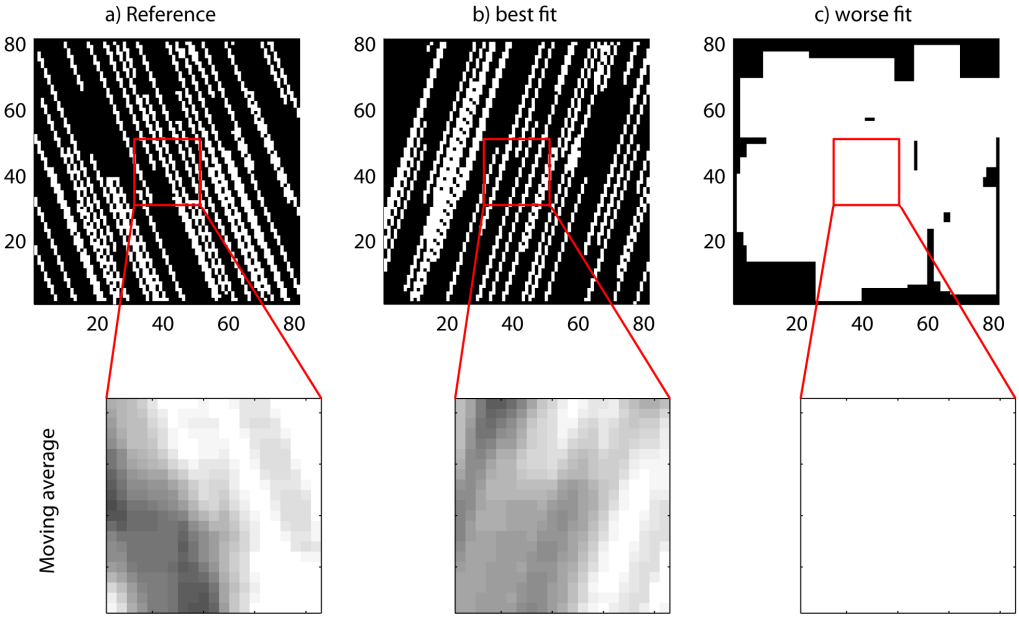

Figure 2. Problem setting. a) Reference realization with data zone and moving average on data zone. b) Best prior realization. c) Worse prior realization.

As stated above, individual parameters in $\boldsymbol{\theta}$ present complex interactions. This is illustrated in Figure 3, where we show the one-way correlation ρ between the error observed for each of the 100 prior models and the parameters used to generate them. No correlation can be identified when parameters are considered individually. This emphasizes the necessity to consider combinations of parameters that carry richer dependencies with the model response.

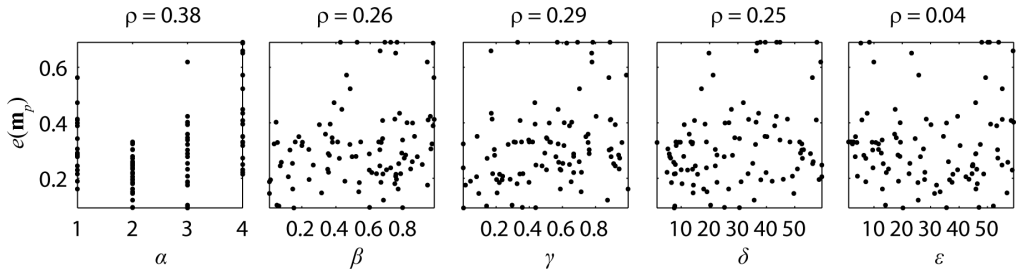

Figure 3. One-way correlation between individual parameters and error towards data reproduction.

3.2. Application and results

We define a distance between the prior parameters vectors using eq. 1 and 2 with the normalization factors η given in Table 1. Using MDS, we obtain a set of points $\mathbf{x}_p$ representing these parameters combinations in $\Re$.

100 models $m(\boldsymbol{\theta}_p)$ are generated. The error associated to these prior models e_p is computed and the likelihood of each model (therefore of each combination of parameters) is modeled as a Gaussian function of e_p having the form

$$L(\mathbf{x}_p) = L(e_p) = \exp\left(-\frac{e_p^{\,2}}{2\sigma^2}\right), \tag{12}$$

with σ = 0.05. Q=200 new samples $\mathbf{y}_q$ are drawn from the posterior using the Metropolis sampler described above, considering 4 dimensions in $\Re$ ind σ_k = 0.2.

Note that σ_k defines the kernel for interpolating $f(\mathbf{x}_p|\mathbf{d})$ in $\Re$, and has to be distinguished from σ which represents the error associated to the data measurements. While a value for σ can be estimated based on measurements errors ranges, choosing an appropriate kernel standard variation σ_k is not straightforward because it represents the smoothness of the likelihood in $\Re$. Ideally, kernels should be large enough to overlap, providing a smooth distribution, but still small enough to distinguish specific favorable areas of $f(\mathbf{x}_p|\mathbf{d})$. As a rule of thumb, we propose to choose σ_k such that it corresponds to half the correlation length of $L(\mathbf{x}_p)$ in order to preserve the local features of the posterior, but large enough to keep continuity. In itself, the variogram $\gamma\{L(\mathbf{x}_p)\}$ (Figure 4) is not very informative since $L(\mathbf{x}_p)$ is likely to be highly non-stationary. However, it gives an order of magnitude of the correlation length that seems to be about 0.4. We also used visual inspection of the results of kernel smoothing (Figure 5a) to choose the value σ_k = 0.2.

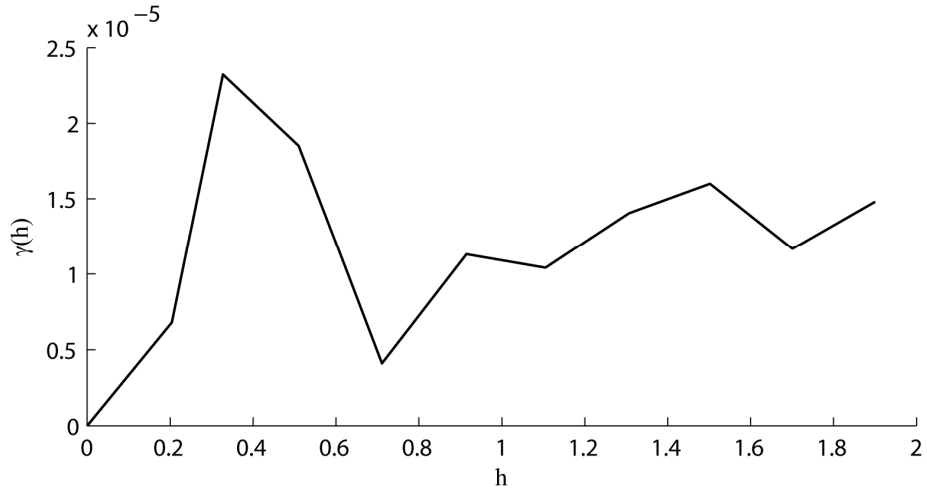

Figure 4. Experimental isotropic variogram of $L(\mathbf{x}_p)$ in $\mathfrak{R}$ (only 4 dimensions are retained).

Figure 5 illustrates the sampling procedure. Figure 5a is the 2D projection of the MDS space $\mathfrak{R}$, where each black dot is a point $\mathbf{x}_p$ representing a combination of parameters $\boldsymbol{\theta}_p$. The background color represents the prior density of points $f(\boldsymbol{\theta}_p)$ obtained by kernel smoothing. Visual inspection of Figure 5a shows that the chosen kernel size of $\sigma_k = 0.2$ provides a continuous probability density function, representative of the density of points.

Figure 5b is a similar representation of the posterior distribution of parameters $f(\boldsymbol{\theta}_p|\mathbf{d})$. The black dots are the same prior parameters $\boldsymbol{\theta}_p$, but here the background color represents the posterior probability obtained with eq. 8. One can observe that areas of high prior density are not necessarily the most likely. In this example, a 2D projection is able to identify two areas of higher likelihood. Magenta dots represent the 200 samples $\mathbf{y}_p$ drawn from $f(\mathbf{x}_p|\mathbf{d})$. Two main clusters are identified in the posterior distribution, and two corresponding realizations are represented (4 dimensions have been used for Metropolis).

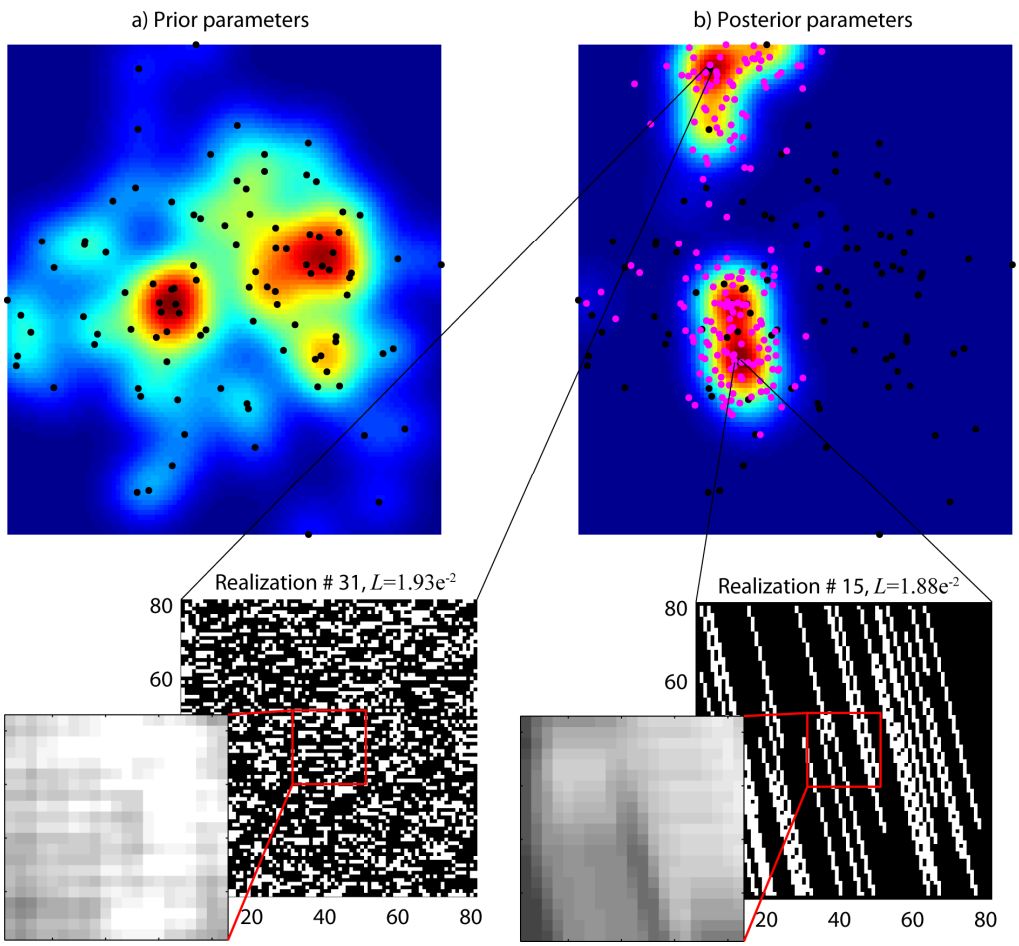

Figure 5. Distance-based representation of the combinations of parameters (2D projection). Axis are not represented because distance spaces are in relative coordinates. a) Prior parameters. b) posterior parameters.

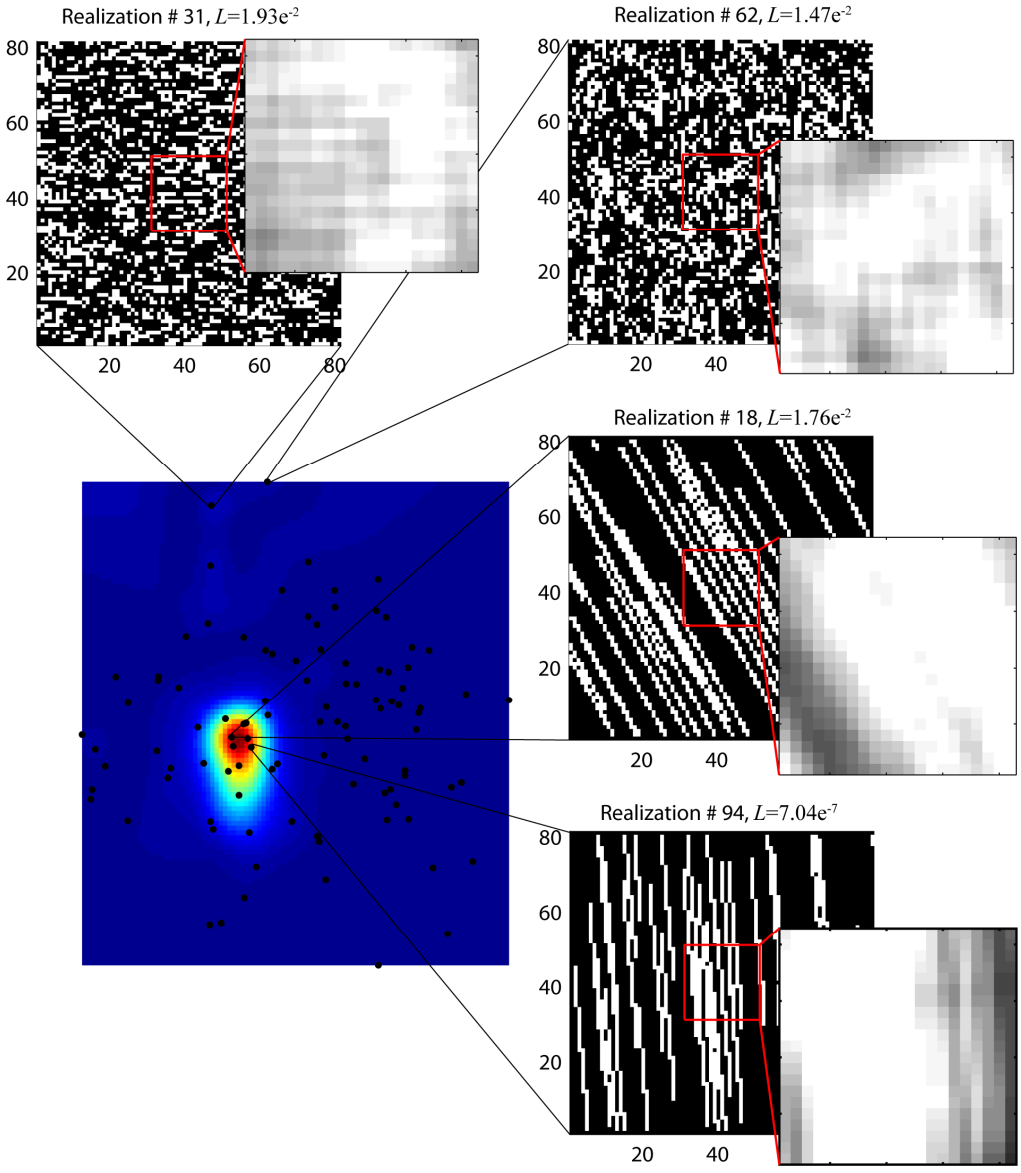

Figure 6. Weighted variance of the likelihood. The likelihood of each realization is displayed in the images titles.

Figure 6 represents a similar 2D mapping for the weighted variance of the likelihood, computed using eq. 9. One can observe large differences for the areas corresponding to the two clusters of Figure 5b. The bottom cluster shows a high variance, meaning that the corresponding models have very different likelihoods. Conversely, the top cluster is made of models that tend to have similar likelihoods, indicating a smaller spatial uncertainty. This difference can be explained by

inspecting the characteristics of the models belonging to each cluster (two realizations of each cluster are represented on Figure 6). In the bottom cluster, realizations are made of lines. Slightly shifting the lines can result in very different results in the data zone, and therefore yields a wide range of likelihoods (the two realizations represented have likelihoods of $7.04e^{-7}$ and $1.76e^{-2}$). Spatial uncertainty is therefore critically contributing to the data misfit in this area. On the contrary, models of the top cluster are made of evenly distributed small patches that result in similar images when smoothed. This explains why the likelihood is more uniform in this area ($1.47e^{-2}$ and $1.76e^{-2}$ for the two realizations represented), corresponding to a low spatial uncertainty. Hence producing additional realizations with the parameters of this cluster is unlikely to yield better fits to the data. This analysis indicates that in one group a large number of models should be generated and evaluated to capture spatial uncertainty. In the other group, it may be sufficient to evaluate a limited number of models since spatial uncertainty is not significant.

By solving a pre-image problem, 200 parameters sets $\boldsymbol{\theta}_q$ are obtained from $\mathbf{y}_p$. These parameters are in turn used to generate a new set of models $m(\boldsymbol{\theta}_q)$. Figure 7 compares the 200 posterior models and the 100 prior models in terms of likelihood (Figure 7a) and in terms of error (Figure 7b). For the present example, the mean likelihood of the posterior models is 4.5 times larger than for the prior models, hence our method allows efficiently narrowing the range of parameters used by the algorithm that produces geological models.

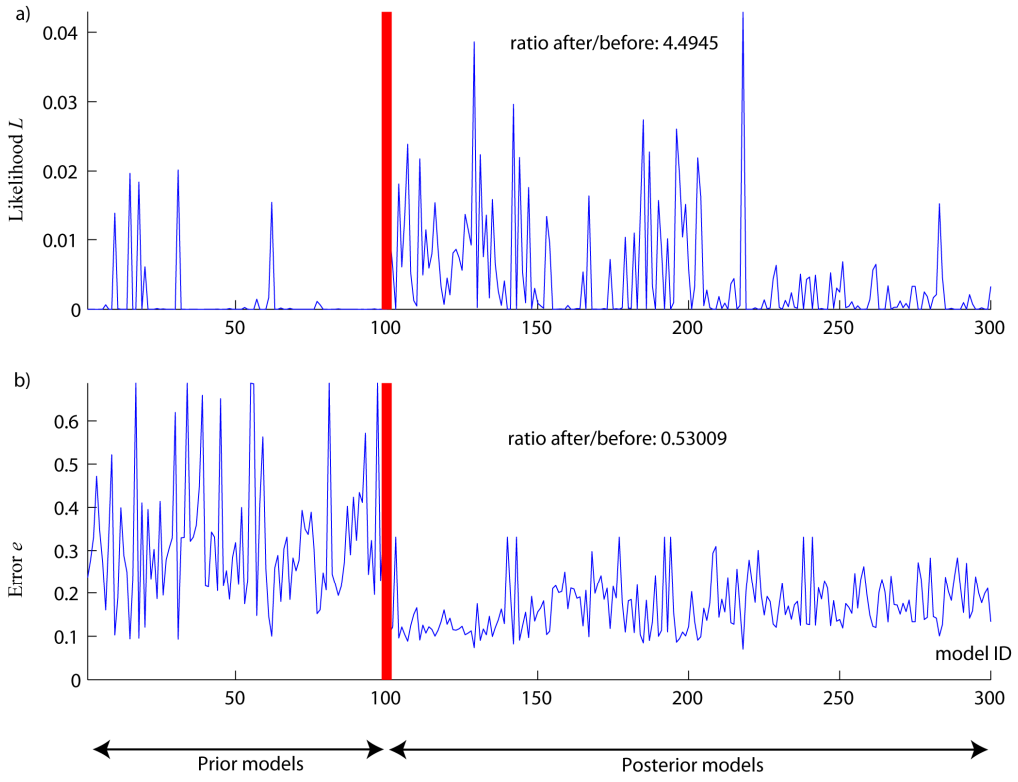

Figure 7. Comparison between prior and posterior models in terms of a) likelihood and b) error.

It is possible to map prior and posterior models in a distance space $\wp$ using MDS with the model error e_p as a distance between models. It is also possible to represent the location of the data in such a space [*Scheidt and Caers*, 2009a]. In $\wp$, all models $m(\mathbf{\theta}_p)$ and $m(\mathbf{\theta}_q)$ are represented as points. Figure 8 shows the 2D projection of $\wp$. Prior models are in black and posterior models in magenta. The data $\mathbf{d}_{ref}$ are represented as the large green dot. It is clear that prior models are spread over a larger area, while the posterior models are more focused around the data.

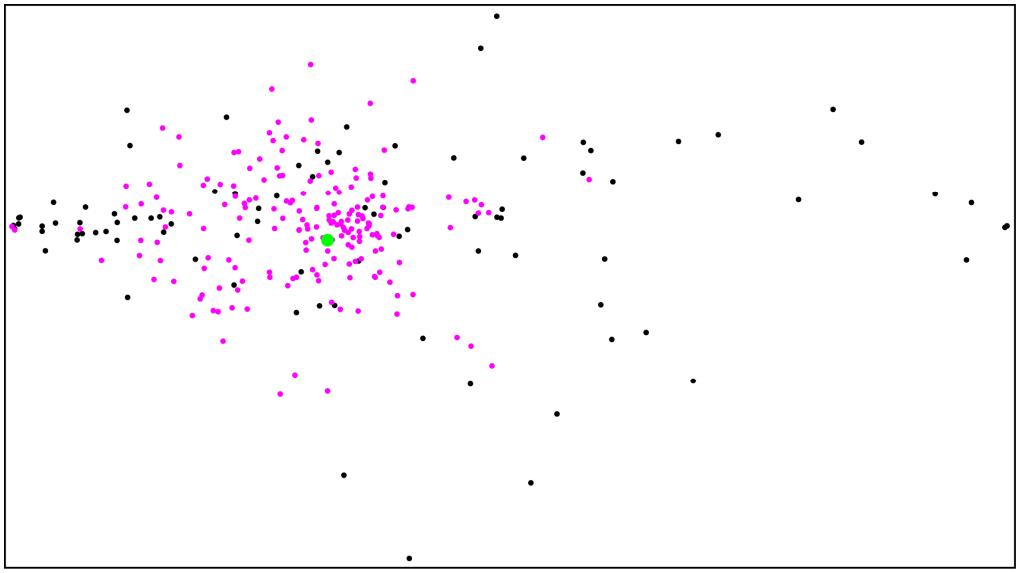

Figure 8. Distance-based representation of the ensembles of models, with a distance considering the error to the data (2D projection). Axis are not represented because distance spaces are in relative coordinates.

Finding the most relevant combinations of parameters (or parameters envelopes) is then a matter of identifying the combinations that yield the best match to the data. Figure 9a displays the combinations of parameters that are associated to the 10% best models (red lines). The blue line is the combination of parameters used to generate the reference $\mathbf{\theta}_{ref}$. One can identify two groups of parameters that share similar characteristics, and that reflect the two clusters observed in the posterior parameters (Figure 5b). The first group (Figure 9b) is similar to $\mathbf{\theta}_{ref}$, which consists in objects of type 2 (lines) with similar parameters as the reference. It corresponds to the bottom

cluster of Figure 5b. The second group consists in truncated Gaussian realizations that have short correlation lengths in both X and Y directions (β, δ and γ are small), and corresponds to the top cluster of Figure 5b.

This example shows that our methodology can effectively provide new updated parameters combinations, but also offers means of identifying and interpreting the significance of these new parameter combinations. Moreover, it allows distinguishing whether spatial uncertainty or parameters uncertainty dominates the variability of model responses.

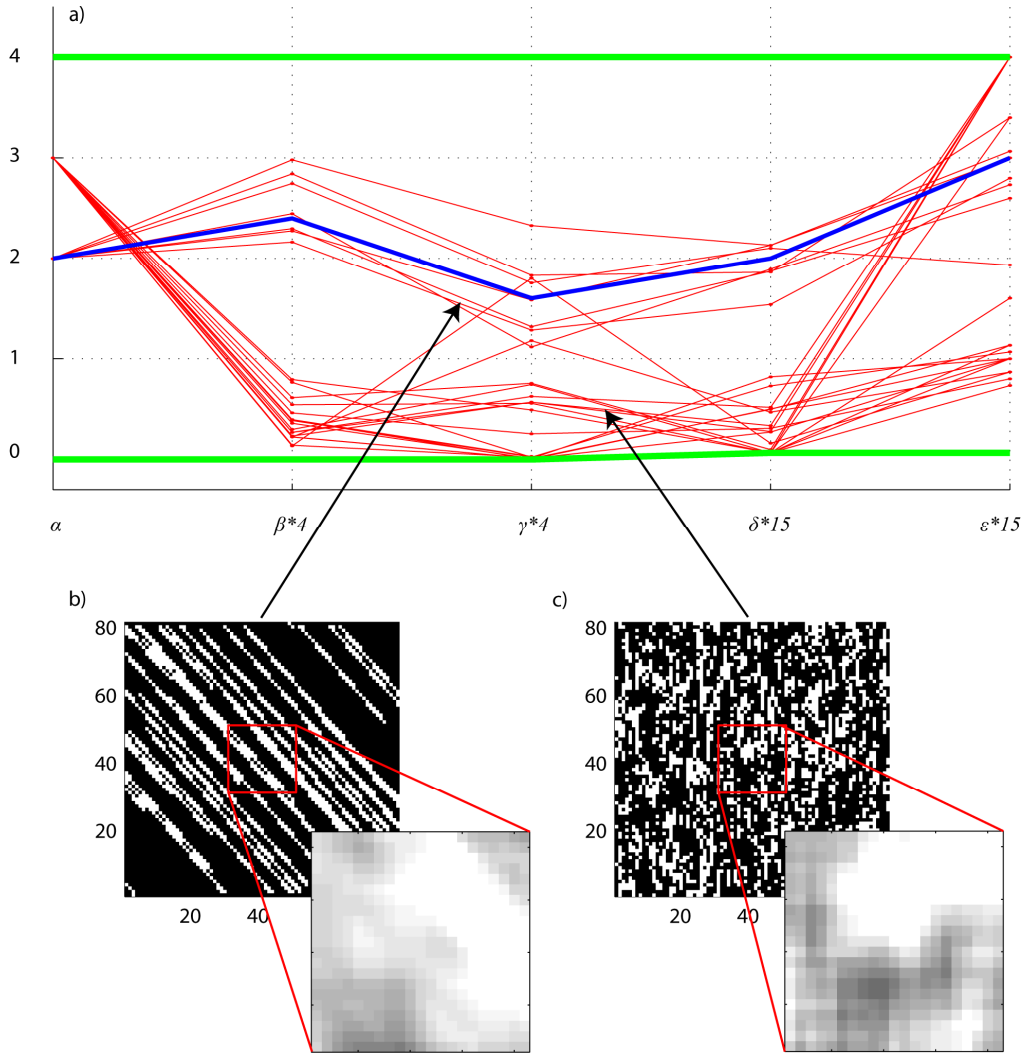

Figure 9. Parameters envelopes for the 10% best posterior models.

4. Conclusions

The methodology presented in this paper allows defining better the parameters that control geological continuity in subsurface models. Although illustrated with object-based simulation, it is general and can be applied to any geostatistical simulation technique. It consists in updating some prior parameters to obtain a posterior parameters distribution. This procedure allows excluding a vast majority of unlikely parameters combinations. The models generated using updated parameters, although not directly matched to the data, honor them better because they belong to a smaller and more focused ensemble. The use of distance-based techniques offers the flexibility to consider combinations of parameters rather than adjusting one parameter at the time.

Increased focus in the parameters range can be highly beneficial for history matching techniques. With focused parameters, the search for history matched models is undergone in a space of reduced dimension, mitigating problems of ill-posedness. Analyzing the posterior parameters distribution allows interpreting the significance of the relevant combinations of parameters. In addition, by using a likelihood weighted variance we can distinguish spatial uncertainty from parameters uncertainty.

In the example presented, it is possible to identify two groups of models parameters that have the potential to better match the data. One group of parameters combinations systematically yields models of high likelihood, while the other group produces realizations whose likelihood can vary by several orders of magnitude. Such an analysis can help determining the most appropriate way of dealing with uncertainty. If one or the other type of uncertainty dominates a given problem, it could be more appropriate to randomize the parameters (e.g. the objects types) or the spatial characteristics (the location of the objects).

Our method would be inefficient in cases where the likelihood of the models is independent of the model parameters. In this case, the posterior parameters distribution is identical to the prior parameters distribution. Then, no updating of the posterior takes place and subsequent models are drawn from the prior. Although inefficient with this worst case scenario, our method would not introduce bias in the sampling since the prior distribution is still honored. Conversely, if some combinations of parameters are more favorable than others, our method allows restricting the search only to the combinations of parameters that matter.

References

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., and T. Hoffman (2006), The probability perturbation method: A new look at Bayesian inverse modeling, *Math. Geol.*, *38*(1), 81-100.
Deutsch, C., and T. Tran (2002), FLUVSIM: a program for object-based stochastic modeling of fluvial depositional systems, *Computers & Geosciences*, *2002*(28), 525-535.
Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170-177.
Epanechnikov, V. A. (1969), Nonparametric estimation of a multidimensional probability density, *Theoretical Probability Applications*(14), 153-158.
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.
Guardiano, F., and M. Srivastava (1993), Multivariate geostatistics: Beyond bivariate moments, in *Geostatistics-Troia*, edited by A. Soares, pp. 133-144, Kluwier Academic, Dordrecht.
Hu, L. (2000), Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models, *Mathematical Geology*, *32*(1), 87-108.
Journel, A., and T. Zhang (2006), The Necessity of a Multiple-Point Prior Model, *Mathematical Geology*, *38*(5), 591-610.
Kitanidis, P. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411-2419.
Lagarias, J., et al. (1998), Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions, *SIAM Journal of Optimization*, *9*(1), 112-147.
Lantuéjoul, C. (2002), *Geostatistical simulation: Models and algorithms*, 232 pp., Springer, Berlin.
Mariethoz, G., and P. Renard (2010), Reconstruction of incomplete data sets or images using Direct Sampling, *Mathematical Geosciences*, *42*(3), 245-268.
Mariethoz, G., et al. (submitted), Bayesian inverse problem and optimization with Iterative Spatial Resampling, *Water Resour. Res.*
Michael, H., et al. (in press), Combining geologic-process models and geostatistics for conditional simulation of 3-D subsurface heterogeneity, *Water Resour. Res.*
Mika, S., et al. (1999), Kernel PCA and de-noising in feature spaces, paper presented at Proceedings of the 1998 conference on Advances in neural information processing systems II, MIT Press, San Mateo, CA.
Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *Journal of geophysical Research*, *100*(B7), 12,431-412,447.
Nelder, A., and R. Mead (1965), A Simplex method for function minimization, *The Computer Journal*, *7*(4), 308-313.
Omre, H., and H. Tjelmland (1996), Petroleum geostatistics, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
Scheidt, C., and J. Caers (2009a), Representing Spatial Uncertainty Using Distances and Kernels, *Mathematical Geosciences*, *41*(2009), 397-419.
Scheidt, C., and J. Caers (2009b), Uncertainty Quantification in Reservoir Performance Using Distances and Kernel Methods—Application to a West Africa

Deepwater Turbidite Reservoir, *SPE Journal*, *14*(4), 680-692.
Scheidt, C., and J. Caers (2010), Bootstrap confidence intervals for reservoir model selection techniques, *Computational Geosciences*, *14*, 369-382.
Suzuki, S., and J. Caers (2008), A Distance-based Prior Model Parameterization for Constraining Solutions of Spatial Inverse Problems, *Mathematical Geosciences*, *40*(2008), 445-469.
Tarantola, A. (2005), *Inverse Problem Theory and Methods for Parameter estimation*, Society for Industrial and Applied Mathematics, Philadelphia.
Zhang, X., et al. (2009), Stochastic surface modeling of deepwater depositional systems for improved reservoir models, *Journal of Petroleum Science and Engineering*, *68*, 118-134.