Mathematical Geology, Vol. 38, No. 1, January 2006 (© 2006)
DOI: 10.1007/s11004-005-9005-9

The Probability Perturbation Method: A New Look at Bayesian Inverse Modeling[1]

Jef Caers[2] and Todd Hoffman[2]

Building of models in the Earth Sciences often requires the solution of an inverse problem: some unknown model parameters need to be calibrated with actual measurements. In most cases, the set of measurements cannot completely and uniquely determine the model parameters; hence multiple models can describe the same data set. Bayesian inverse theory provides a framework for solving this problem. Bayesian methods rely on the fact that the conditional probability of the model parameters given the data (the posterior) is proportional to the likelihood of observing the data and a prior belief expressed as a prior distribution of the model parameters. In case the prior distribution is not Gaussian and the relation between data and parameters (forward model) is strongly non-linear, one has to resort to iterative samplers, often Markov chain Monte Carlo methods, for generating samples that fit the data likelihood and reflect the prior model statistics. While theoretically sound, such methods can be slow to converge, and are often impractical when the forward model is CPU demanding. In this paper, we propose a new sampling method that allows to sample from a variety of priors and condition model parameters to a variety of data types. The method does not rely on the traditional Bayesian decomposition of posterior into likelihood and prior, instead it uses so-called pre-posterior distributions, i.e. the probability of the model parameters given some subset of the data. The use of pre-posterior allows to decompose the data into so-called, "easy data" (or linear data) and "difficult data" (or nonlinear data). The method relies on fast non-iterative sequential simulation to generate model realizations. The difficult data is matched by perturbing an initial realization using a perturbation mechanism termed "probability perturbation." The probability perturbation method moves the initial guess closer to matching the difficult data, while maintaining the prior model statistics and the conditioning to the linear data. Several examples are used to illustrate the properties of this method.

KEY WORDS: geostatistics; inverse modeling; Bayes' theory; history matching.

INTRODUCTION

Inverse problems are ubiquitous in the Earth Sciences. Sets of measurements **d** are used to determine the spatial distribution of a physical attribute, described mathematically by a model with a set of parameters **m**. In most applications, **d** is too sparse to determine uniquely the underlying subsurface phenomenon or

[1]Received 20 July 2004; accepted 18 April 2005; Published online: 18 April 2006.
[2]Department of Petroleum Engineering, Montana Tech of the University of Montana, Butte, Montana, 59701; e-mail: thoffman@mtech.edu.

the model **m**. For example, a well-test pressure response (**d**) does not uniquely determine subsurface permeability at every location (**m**) unless the reservoir is purely homogeneous (single m). A set of geophysical measurements such as seismic data does not uniquely determine the subsurface porosity distribution. Most inverse problems are therefore fundamentally underdetermined (ill-posed), with many alternative solutions fitting the same data **d**.

For this reason, a stochastic approach to solving the inverse problem is taken. Instead of determining a single solution, one generates a set of solutions distributed according to a probability distribution function. More specifically, in a Bayesian context one aims at sampling such solutions from the posterior density distribution of the model parameters **m** given the data **d**,

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

The prior density $f(\mathbf{m})$ describes the dependency between the model parameters and therefore constrains the set of possible inverse solutions. In a spatial context, such dependency refers to the spatial structure of **m**, be it a covariance, Boolean or training image-defined dependency. The likelihood density $f(\mathbf{d}|\mathbf{m})$ models the stochastic relationship between the observed data **d** and each particular model **m** retained. This likelihood would account for model and measurement errors. Void of any such errors, the data and model parameters are related through a forward model g,

$$\mathbf{d} = g(\mathbf{m}) \tag{2}$$

The density $f(\mathbf{d})$ depends on the density $f(\mathbf{m})$ and the forward model g, but its specification can often be avoided in most samplers of $f(\mathbf{m}|\mathbf{d})$.

The aim of Bayesian inverse methods is to obtain samples of the posterior density $f(\mathbf{m}|\mathbf{d})$. Other than the case where g is linear, such sampling will need to be iterative. Popular sampling methods are rejection sampling and the Metropolis sampler (Besag and Green, 1993; Metropolis and others, 1953; Mosegaard and Tarantola, 1995; Neal, 1993; Omre and Tjelmeland, 1996). Both are Markov chain Monte Carlo (McMC) samplers that avoid the specification of $f(\mathbf{d})$ but are iterative in nature in order to obtain a single sample $\mathbf{m}^{(\ell)}$ of $f(\mathbf{m}|\mathbf{d})$. Generating multiple (conditioned to **d**) samples $\mathbf{m}^{(\ell)}$, $\ell = 1, \ldots, \mathrm{L}$ in this manner quantifies the uncertainty modeled in $f(\mathbf{m}|\mathbf{d})$.

Some practical limitations of McMC approaches warrant further discussion:

1. Iterative samplers may require many thousands of evaluations of the forward model g. In practical cases, e.g., flow modeling and ray tracing in a complex 3D subsurface, a single evaluation of the forward model may require several hours of CPU time.

2. The multi-Gaussian distribution is the only case for which the decomposition (1) is analytically known (likelihood, posterior and priors are all multi-Gaussian if g is a linear function or can be sufficiently linearized). For this mathematical convenience, the multi-Gaussian is popular in Bayesian inverse work (e.g., Hegstad and Omre, 1996; Moosegard and Tarantola, 1995; Tarantola, 1987). The Gaussian framework requires a covariance structure for the prior as well as likelihood. The prior covariance is estimated from any direct observations of the model parameters $\mathbf{m}$. The covariance structure of the likelihood often has to be assumed. Homoscedasticity of the error distribution, namely independence of error ε with regard to the "signal" $g(\mathbf{m})$ in

$$\mathbf{d} = g(\mathbf{m}) + \varepsilon$$

is assumed, to make inference of the likelihood distribution feasible.

The Gaussian framework is too limited to capture the possibly large variety of prior distributions of $\mathbf{m}$ observed in reality, neither does it properly model the full dependency between $\mathbf{m}$ and $\mathbf{d}$ in the likelihoods. Most spatial phenomena exhibit a strong correlation in the form of non-rectilinear shapes and connectivity, e.g. channels, cross—bedding, complex litho-facies distributions that cannot be adequately described by making multi-Gaussian assumptions.

In this paper, we also treat the more general case where two types of dataset $\mathbf{d}_1$ and $\mathbf{d}_2$ are available. In such case, the Bayesian decomposition of (1) is written as

$$\begin{aligned} f(\mathbf{m}|\mathbf{d}_1, \mathbf{d}_2) &= \frac{f(\mathbf{d}_1, \mathbf{d}_2|\mathbf{m}) f(\mathbf{m})}{f(\mathbf{d}_1, \mathbf{d}_2)} \\ &\sim \frac{f(\mathbf{d}_1|\mathbf{m}) f(\mathbf{d}_2|\mathbf{m}) f(\mathbf{m})}{f(\mathbf{d}_1, \mathbf{d}_2)} \end{aligned} \tag{3}$$

where the latter expression relies on a conditional independence hypothesis between the two data types. We will consider the data $\mathbf{d}_1$ as the "easy data," i.e. data can then be conditioned to within a more classical geostatistical context. The data $\mathbf{d}_1$ could consist of a set of point measurements (termed "hard data") combined with secondary information (termed "soft data") whose relationship to the model $\mathbf{m}$ is linear or pseudo-linear in nature. The data $\mathbf{d}_2$ is termed the "difficult data," exhibiting a complex multi-point and non-linear relationship with the model $\mathbf{m}$.

In this paper, we present a practical method, termed *probability perturbation*, that addresses the above—mentioned limitations. The method relies on

1. The use of sequential simulation (Deutsch and Journel, 1998) to sample the prior and posterior distributions. Sequential simulation is not iterative, hence CPU efficient.

2. The use of so-called pre-posterior distributions $f(\mathbf{m}|\mathbf{d}_1)$ and $f(\mathbf{m}|\mathbf{d}_2)$, instead of likelihoods. We will show that this leads to an efficient sampling method, primarily by avoiding the calculation of $f(\mathbf{d}_1, \mathbf{d}_2)$.

The probability perturbation method has been previously published as an algorithm in field-specific papers (Caers, 2003) on history matching (inversion of flow and pressure data), including a large case study (Hoffman and Caers, 2004). A concise theoretical framework however was still lacking. The novelty in this paper lies in framing this method within the traditional Bayesian inverse context and in demonstrating the theoretical foundation of the approach. For sake of demonstration, we present an example of an inverse problems that deals with subsurface flow, although as will be shown, the method has the potential to address other inverse problems.

METHODOLOGY

Sampling the Prior

Practical inverse methods should be able to incorporate a diversity of prior models, not limited to Gaussian. To emphasize the non-Gaussianity of the proposed method, we consider a binary random function, modeling the absence or presence of a geological event,

$$I(\mathbf{u}) = \begin{cases} 1, & \text{if the "event" occurs at } \mathbf{u} \\ 0, & \textit{else} \end{cases}$$

where "event" could represent any spatially distributed, whether continuous or categorical, phenomenon, such as lithologies, petrophysical, or geophysical properties. The random function is discretized on a grid composed of a finite set of N grid node locations $\mathbf{u}_i = (x_i,y_i,z_i)$. The parameters of the inverse problem are given by a set of indicator variables at each grid node location:

$$\mathbf{m} = \{I(\mathbf{u}_1), I(\mathbf{u}_2), \ldots, I(\mathbf{u}_N)\}$$

The grid need not be rectangular. The prior distribution is then simply the joint distribution of all indicator random variables at all grid node locations

$$f(\mathbf{m}) = \text{Prob}\{I(\mathbf{u}_1) = i(\mathbf{u}_1), I(\mathbf{u}_2) = i(\mathbf{u}_2), \ldots, I(\mathbf{u}_N) = i(\mathbf{u}_N)\}$$

A fast and general method for sampling a prior distribution is sequential simulation. Unlike McMC methods, sequential simulation is non-iterative and is completed

after a single pass over all grid node locations. Sequential simulation relies on the following decomposition of the prior:

$$\begin{aligned} f(\mathbf{m}) = \text{Prob}\{I(\mathbf{u}_1) = 1\} \times \\ \text{Prob}\{I(\mathbf{u}_2) = 1 | i(\mathbf{u}_1)\} \times \cdots \times \\ \text{Prob}\{I(\mathbf{u}_N) = 1 | i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{N-1})\} \end{aligned} \tag{4}$$

This decomposition can be interpreted in two ways:

1. Sampling a given joint distribution $f(\mathbf{m})$ is equivalent to sequentially sampling a series of univariate conditional distributions. If the joint distribution is known analytically, it determines the shape of each conditional distribution. The only joint distribution for which each conditional is known is the multi-variate Gaussian. In case of a joint multi-Gaussian, each conditional is also Gaussian with mean and variance determined by a kriging system. Sequential Gaussian simulation (sgsim) is one of the most efficient algorithms for drawing samples from the multi-Gaussian distribution.
2. Sequential drawing from a series of given univariate conditional distributions provides a realization (sample) of joint distribution which need not be explicitly stated. The only restriction on each such univariate distribution is that it needs to be conditional on all previously simulated parameters of the vector $\mathbf{m}$. The aim in this approach is to generate realizations of $\mathbf{m}$ exhibiting certain (univariate, bivariate, or multivariate) statistics. The set of realizations generated in this manner represents a sample from some, not explicitly known, joint distribution.

The second approach recognizes that one is not interested in $f(\mathbf{m})$ itself but in the realizations generated from $f(\mathbf{m})$. In actual field cases, prior information on $\mathbf{m}$ comes in the form of limited statistics (e.g. a spatial covariance). The type of multi-variate density f cannot be determined from data and always needs to be assumed. In the second approach, the decision of distribution type is not made on the joint distribution, but on the conditional distributions. An example of such approach is direct sequential simulation (dssim, Journel, 1993), where the conditional distribution can be of any type, as long as they have mean and variance provided by a simple kriging system.

In practice, we obtain multiple realizations of $f(\mathbf{m})$ by changing the random seed used to generate stochastic simulations. The random seed determines the order in which grid nodes are visited (random path) as well as the uniform numbers that are used to draw from each conditional density function.

The second interpretation has led to a series of practical algorithms that can sample a wide variety of prior models. One such method that will be briefly

reviewed is the *single normal equation simulation* or *snesim* in short (Journel, 1993; Strebelle, 2002).

In most geostatistical applications, the variogram (or spatial covariance) is used as a quantifier for spatial heterogeneity. However, the variogram, as a two-point statistic, cannot capture non-rectilinear shapes and connectivities which require the use of higher order, also termed multi-point statistics. Such higher order statistics can rarely be estimated from actual field data. Instead, one relies on the construction of a so-called training images. The training image is a fully explicit (2D or 3D) depiction of the spatial heterogeneity present. It is merely conceptual, hence has to be constrained to any data ($\mathbf{d}_1$ or $\mathbf{d}_2$).

The snesim algorithm (Strebelle, 2002) is a practical sequential simulation method for generating stochastic realization depicting the structure of the training image. At each visit of a location $\mathbf{u}$ along the random path, the snesim method calculates the conditional probability $\text{Prob}\{I(\mathbf{u}_j) = 1 | i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1})\}$ by scanning the training image for replicates of the joint event $\mathbf{d}_j^{(1)} = \{i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1}), I(\mathbf{u}_j) = 1\}$ and $\mathbf{d}_j^{(0)} = \{i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1}), I(\mathbf{u}_j) = 0\}$. The conditional probability is then calculated as follows:

$$\begin{aligned} &\text{Prob}\{I(\mathbf{u}_j) = 1 | i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1})\} \\ &\quad = \frac{\text{Prob}\{I(\mathbf{u}_j) = 1, I(\mathbf{u}_1) = i(\mathbf{u}_1), \ldots, I(\mathbf{u}_{j-1}) = i(\mathbf{u}_{j-1})\}}{\text{Prob}\{I(\mathbf{u}_1) = i(\mathbf{u}_1), \ldots, I(\mathbf{u}_{j-1}) = i(\mathbf{u}_{j-1})\}} \\ &\quad \sim \frac{\#\text{ of events } \mathbf{d}_j^{(1)}}{\#\text{ of events } \mathbf{d}_j^{(0)} + \#\text{ of events } \mathbf{d}_j^{(1)}} \end{aligned} \tag{5}$$

Bayesian Inversion

Priors need to be conditioned to data $\mathbf{d}_1$ and $\mathbf{d}_2$, resulting in a posterior distribution of the model parameters given the data, as formulated in Bayes rule, Equations (1) or (3). In determining the posterior, we will follow the same philosophy as in determining the prior: the posterior will not be explicitly stated, only conditionals will be modeled. We consider the case where the data $\mathbf{d}_1$ constitutes direct observations (hard data) of the model parameters at a set of $n < N$ spatial locations, in the binary random function case,

$$\mathbf{d}_1 = \{i(\mathbf{u}_\alpha),\ \alpha = 1, \ldots, n\}$$

$\mathbf{d}_2$ contains any type of data that has a non-linear relationship with the model parameters

$$\mathbf{d}_2 = g(\mathbf{m}) = g(I(\mathbf{u}_1), I(\mathbf{u}_2), \ldots, I(\mathbf{u}_N))$$

with g as the forward model. We will assume for now that there are no data and model errors. One is interested in generating samples from a posterior distribution:

$$f(\mathbf{m}|\mathbf{d}_1, \mathbf{d}_2) = \text{Prob}\{I(\mathbf{u}_1) = i(\mathbf{u}_1),\ I(\mathbf{u}_2) = i(\mathbf{u}_2), \ldots, I(\mathbf{u}_N) = i(\mathbf{u}_N)|\{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}, \mathbf{d}_2\}$$

To sample such posterior, we do not follow the traditional Bayesian route of likelihood and prior. Instead, we rely on the sequential decomposition similar to the one in the prior case:

$$\begin{aligned} f(\mathbf{m}|\mathbf{d}_1, \mathbf{d}_2) = {} & \text{Prob}\{I(\mathbf{u}_1) = 1|\{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}, \mathbf{d}_2\} \times \\ & \text{Prob}\{I(\mathbf{u}_2) = 1|i(\mathbf{u}_1), \{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}, \mathbf{d}_2\} \times \cdots \times \\ & \text{Prob}\{I(\mathbf{u}_N) = 1|i(\mathbf{u}_1), \ldots, i(u_{N-1}), \{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}, \mathbf{d}_2\} \end{aligned} \tag{6}$$

Sampling from the complex (joint) posterior f is equivalent to sequential sampling from a series of univariate conditionals of the type

$$\begin{aligned} &\text{Prob}\{I(\mathbf{u}_j) = 1|i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1}), \{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}, \mathbf{d}_2\} = \text{Prob}(A_j|\mathbf{B}_j, \mathbf{C}) \\ &\quad \text{with } A_j = \{I(\mathbf{u}_j) = 1\} \\ &\qquad \mathbf{B}_j = \{i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1}), \{i(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}\} \\ &\qquad \mathbf{C} = \mathbf{d}_2 \end{aligned}$$

where we have introduced a simpler notation A_j (unknown model parameter), $\mathbf{B}_j$ data (easy data) and $\mathbf{C}$ data (difficult data) to make further development clear. Since it is difficult to state the conditionals $\text{Prob}(A_j|\mathbf{B}_j, \mathbf{C})$ explicitly, we decompose it further into two "pre-posterior" distributions, $\text{Prob}(\mathbf{A}_j|\mathbf{B}_j)$ (pre-posterior to knowing $\mathbf{C}$) and $\text{Prob}(A_j|\mathbf{C})$ (pre-posterior to knowing $\mathbf{B}_j$), using Journel's (2002) combination method called "tau-model":

$$\text{Prob}(A_j|\mathbf{B}_j, \mathbf{C}) = \frac{1}{1+x} \quad \text{with} \quad \frac{x}{a} = \left(\frac{b}{a}\right)^{\tau_1} \left(\frac{c}{a}\right)^{\tau_2} \tag{7}$$

where

$$b = \frac{1 - \text{Prob}(A_j|\mathbf{B}_j)}{\text{Prob}(A_j|\mathbf{B}_j)}, \qquad c = \frac{1 - \text{Prob}(A_j|\mathbf{C})}{\text{Prob}(A_j|\mathbf{C})}, \qquad a = \frac{1 - \text{Prob}(A_j)}{\text{Prob}(A_j)}$$

In case, $\tau_1 = \tau_2 = 1$, Journel shows that Equation (7) is equivalent (up to a simple standardization) to the hypothesis of conditional independence of Equation (3).

Assuming the prior Prob(A_j) is known, the problem of stating the conditional Prob($A_j|\mathbf{B}_j, \mathbf{C}$) is now decomposed into a problem of stating the pre-posteriors Prob($A_j|\mathbf{B}_j$) and Prob($A_j|\mathbf{C}$).

Working with pre-posteriors will lead to an approach that is different from a classical Bayesian inversion which would involve the likelihoods Prob($\mathbf{B}_j|A_j$) and Prob($\mathbf{C}|A_j$). This seemingly subtle difference will lead to a fundamentally different approach. Stating "pre-posteriors," i.e. likelihoods, allows using (non-iterative) sequential simulation via Equations (6) and (7), instead of (iterative) McMC.

The τ-values in Equation (7) allow the user to model explicitly the dependency between the **B**-data and **C**-data. The τ-values can be interpreted as "weights" given to each data type (see Journel, 2002). Assuming conditional independence ($\tau_1 = \tau_2 = 1$) results in a very particular dependency model that may often not reflect the actual dependency between **B** and **C** data (see further for an example study).

In the context of sequential simulation, the pre-posterior Prob($A_j|\mathbf{B}_j$) is simply the conditional distribution of the unknown A_j given any previously simulated nodes $i(\mathbf{u}_1), \ldots, i(\mathbf{u}_{j-1})$ and the hard data $i(\mathbf{u}_\alpha),\ \alpha = 1, \ldots, n$. In *snesim*, Prob($A_j|\mathbf{B}_j$) is calculated from the training image using Equation (5).

The remaining pre-posterior Prob($A_j|\mathbf{C}$) cannot be directly estimated because of the complex non-linear relationship (=forward model) between A_j and **C**. Instead, in the next section, we propose an iterative calibration method for determining this probability, termed probability perturbation.

Sampling Prob($A_j|\mathbf{B}_j, \mathbf{C}$): The Probability Perturbation Method

Given a random seed s, and given the pre-posterior Prob($A_j|\mathbf{B}_j$) at all grid nodes $\mathbf{u}_j$, a realization $\mathbf{i}_B^{(s)} = \{i_B^{(s)}(\mathbf{u}_1), i_B^{(s)}(\mathbf{u}_2), \ldots, i_B^{(s)}(\mathbf{u}_N)\}$ can be drawn by ways of sequential simulation. The subscript B in $\mathbf{i}_B^{(s)}$ emphasizes that $\mathbf{i}_B^{(s)}$is conditioned to data $\mathbf{d}_1$ (**B**-data) only, not yet to data $\mathbf{d}_2$ (**C**-data), this would require somehow the use of the other pre-posterior, Prob($A_j|\mathbf{C}$).

To generate a realization $\mathbf{i}^{(\ell)}$ matching both $\mathbf{d}_1$ and $\mathbf{d}_2$, each grid node should be sequentially sampled from the joint distribution Prob($A_j|\mathbf{B}_j, \mathbf{C}$). However, the joint distribution is unknown, since Prob($A_j|\mathbf{C}$) is not known. The probability perturbation method performs a stochastic search for those probabilities Prob($A_j|\mathbf{C}$) (at all N grid nodes) that achieve a match to the $\mathbf{d}_2$ data.

The key idea is to search for all N probabilities Prob($A_j|\mathbf{C}$), $\forall j$ such that, after combining Prob($A_j|\mathbf{C}$) with Prob($A_j|\mathbf{B}$) into Prob($A_j|\mathbf{B}_j, \mathbf{C}$) using Equation (7), a realization $\mathbf{i}^{(\ell)}$ sequentially drawn from Prob($A_j|\mathbf{B}_j, \mathbf{C}$), $j = 1, \ldots, N$, matches the data $\mathbf{d}_2$. Searching for all N probabilities Prob($A_j|\mathbf{C}$) directly is impossible, particularly since N can be large. Therefore, we rely on a perturbation

parameterization of all N Prob$(A_j|\mathbf{C})$ using a single parameter that is defined as follows:

$$\text{Prob}(A_j|\mathbf{C}) = \text{Prob}(I(\mathbf{u}_j) = 1|\mathbf{C}) = (1 - r_C) \times i_B^{(s)}(\mathbf{u}_j) + r_C P(A_j), \quad j = 1, \ldots, N \tag{8}$$

where r_C is a parameter between [0,1], not dependent on $\mathbf{u}_j$. This independence on $\mathbf{u}_j$ will be relaxed later. Given Equation (8), Prob$(A_j|\mathbf{C})$ can be calculated for a given value of r_C and for a given initial realization $i_B^{(s)}(\mathbf{u}_j)$. The importance of Equation (8) is that it translates the search for N probabilities Prob$(A_j|\mathbf{C})$ into an optimization problem of a *single* parameter r_C as follows:

- Choose a random seed s
- Generate an initial realization $\mathbf{i}_B^{(s)}$ using the data $\mathbf{d}_1$ and random seed s
- *Iterate*: Until the data $\mathbf{d}_2$ is matched, do the following:
  1. Choose another random seed s'
  2. Determine a realization $i^{(s')}$ that matches better data $\mathbf{d}_2$ as follows:
     - From Prob$(A_j|\mathbf{C})$ in Equation (8), Prob$(A_j|\mathbf{B}_j, \mathbf{C})$ can be determined using Equation (7). This allows generating a realization $\mathbf{i}_{r_C}^{(s')} = \{i^{(s')}(\mathbf{u}_1), i^{(s')}(\mathbf{u}_2), \ldots, i^{(s')}(\mathbf{u}_N)\}$ drawn by sequentially sampling from Prob$(A_j|\mathbf{B}_j, \mathbf{C})$, $j = 1, \ldots, N$. This realization is dependent on the value of r_C, as well as a random seed s'.
     - To find an optimal r_C, hence per Equation (8), best Prob$(A_j|\mathbf{C})$, the following objective function is formulated:

$$O(r_C, s') = \left\| \mathbf{d}_2 - g\left(\mathbf{i}_{r_C}^{(s')}\right) \right\| \tag{9}$$

       which measures the distance between the data $\mathbf{d}_2$ and the forward model evaluated on the realization $\mathbf{i}_{r_C}^{(s')}$. The objective function (9) depends on r_C and a fixed random seed s'. A simple one-dimensional optimization can be performed to find the value of r_C that best matches the data $\mathbf{d}_2$.
     - Once an optimal value r_C^{opt} is found, generate a realization $\mathbf{i}_{r_C^{\text{opt}}}^{(s')}$ to be used in Equation (8) during the next iteration step.

The two limit values $r_C = 0$ and $r_C = 1$ clarify the choice of the parameterization in Equation (8):

1. In case $r_C = 0$, then Prob$(A_j|\mathbf{C}) = i_B^{(s)}(\mathbf{u}_j)$, hence per Equation (7) Prob$(A_j|\mathbf{B}_j, \mathbf{C}) = i_B^{(s)}(\mathbf{u}_j)$. Regardless of the random seeds s and s', any realization $\mathbf{i}_{r_C=0}^{(s')} = \mathbf{i}_B^{(s)}$. In other words, $r_C = 0$ entails "no perturbation of $\mathbf{i}_B^{(s)}$".

2. In case $r_C = 1$, then $\text{Prob}(A_j|\mathbf{C}) = P(A_j)$, a simple calculation using Equation (7) shows that $\text{Prob}(A_j|\mathbf{B}_j, \mathbf{C}) = P(A_j|\mathbf{B}_j),\ \forall j$. Since the seed s' is different from the seed s, the realization $\mathbf{i}^{(s')}_{r_C=1} = \mathbf{i}^{(s')}_B$ is equiprobable with $\mathbf{i}^{(s)}_B$. In other words, $r_C = 1$ entails a "maximum perturbation" within the prior model constraints.

A value r_C between (0, 1) will therefore generate a perturbation $\mathbf{i}^{(s')}_{r_C}$ between some initial realization $\mathbf{i}^{(s)}_B$ and another equiprobable realization $\mathbf{i}^{(s')}_B$ both conditioned to the data $\mathbf{d}_1$ and each following the prior model statistics.

The optimization of r_C in (9), has to be repeated for multiple random seeds s' since a single optimization of $O(r_C^{\text{opt}}, s')$ with fixed random seed is likely not to reach satisfactory match to the data $\mathbf{d}_2$.

A Simple Illustrative Example

A simple example is presented to clarify the approach and illustrate various properties of the method in finding inverse solutions. The model consists of a grid with three nodes, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. Each node can be either black, $I(\mathbf{u}) = 1$ or white, $I(\mathbf{u}) = 0$. The model $\mathbf{m}$ is therefore simply

$$\mathbf{m} = \{I(\mathbf{u}_1), I(\mathbf{u}_2), I(\mathbf{u}_3)\}$$

The spatial dependency of this simple 1D model is described by a training image shown in Figure 1. One can extract, by scanning the training image with a 3×1 template, the prior distribution, $f(\mathbf{m})$, of the model parameters $\mathbf{m}$, as shown in Figure 1. To test the probability perturbation method, we consider two data: the first data is a point measurement (B-data, or "easy data") namely, $i(\mathbf{u}_2) = 1$ (a

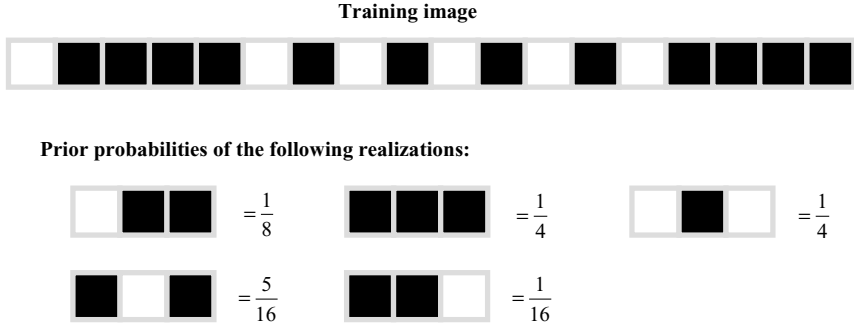

Figure 1. Illustration example: Training image and prior probabilities of a 3×1 model derived from the training image.

black pixel in the middle), the second one is $I(\mathbf{u}_1) + I(\mathbf{u}_2) + I(\mathbf{u}_3) = 2$ (C-data or "difficult data"). The problem posed is:

$$\text{What is Prob}(I(\mathbf{u}_1) = 1 | i(\mathbf{u}_2) = 1, I(\mathbf{u}_1) + I(\mathbf{u}_2) + I(\mathbf{u}_3) = 2)?$$

Or in simpler notation, what is Prob$(A|B, C)$? We consider three ways to solve this problem:

1. Calculate the true posterior Prob$(A|B, C)$ directly by simple elimination of those prior models that do not match the B and C data.
2. Calculate Prob$(A|B, C)$ by first calculating Prob$(A|B)$ and Prob$(A|C)$ based on the prior, then using Journel's Equation (7) under conditional independence, namely $\tau_1 = 1, \tau_2 = 1$.
3. Calculate Prob$(A|B, C)$ by running a Monte Carlo experiment using the probability perturbation method (PPM). A set of realization is generated constrained to the single datum $i(\mathbf{u}_2) = 1$. On each of these realizations, the PPM using the algorithm described above is applied. Figure 2 provides a flow chart for a running the PPM on a single realization. An algorithmic description is as follows:

   i. Set a random seed.
   ii. Perform sequential simulation with conditioning data $i(\mathbf{u}_2) = 1$, all conditional distributions of the type Prob$(A|B_j)$ can be determined from the training image.
   iii. Change seed.

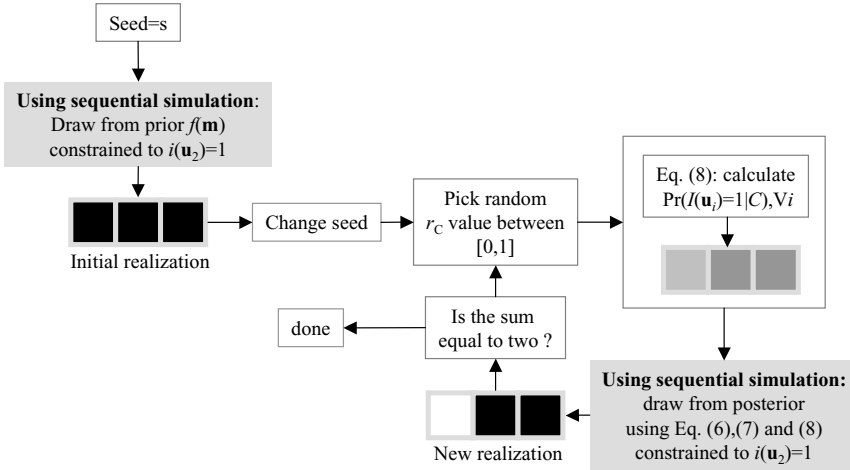

Figure 2. Probability perturbation flowchart: generating multiple realizations constrained to the two data $i(\mathbf{u}_2) = 1$ and $i(\mathbf{u}_1) + i(\mathbf{u}_2) + i(\mathbf{u}_3) = 2$.

iv. Until the sum $i(\mathbf{u}_1) + i(\mathbf{u}_2) + i(\mathbf{u}_3) = 2$ is matched, do the following:

   a. Pick a random value for r_C.
   b. Calculate Prob($I(\mathbf{u}_1) = 1|C$) and Prob($I(\mathbf{u}_3) = 1|C$) from Equation (8).
   c. Perform sequential simulation with conditioning data $i(\mathbf{u}_2) = 1$, all conditional distributions of the type $P(A|B_j,C)$ are determined by combining Prob($A|B_j$) (derived from the training image) and Prob($A|C$) (precalculated using Equation (8)) under conditional independence.

v. If the sum $i(\mathbf{u}_1) + i(\mathbf{u}_2) + i(\mathbf{u}_3)$ does not equal to two then, change seed and goto iv.

From the set of matched realizations, Prob($A|B, C$) is calculated by checking how many times $i(\mathbf{u}_1) = 1$.

4. Calculate Prob($A|B, C$) by running a Monte Carlo experiment using the gradual deformation method (GDM) of sequential simulations (Hu, Blanc, and Noetinger, 2001). A set of realization is generated constrained to the single datum $i(\mathbf{u}_2) = 1$. On each of these realizations, the GDM is applied. This method perturbs not the probability distributions, but the random numbers that are used to generate the conditional simulation. Similar to PPM, GDM parameterized that perturbation using a single parameter that can be optimized.

The solutions are as follows:

1. True posterior: Prob($A|B, C$) $= \frac{1/16}{(1/8)+(1/16)} = \frac{1}{3}$
2. Using conditional independence Equation (7):

$$\text{Prob}(A) = \frac{4}{16} + \frac{1}{16} + \frac{5}{16} = \frac{5}{8} \Rightarrow a = \frac{3}{5}$$

$$\text{Prob}(A|B) = \frac{\frac{1}{4} + \frac{1}{16}}{\frac{1}{4} + \frac{1}{16} + \frac{1}{8} + \frac{1}{4}} = \frac{5}{11} \Rightarrow b = \frac{6}{5}$$

$$\text{Prob}(A|C) = \frac{\frac{5}{16} + \frac{1}{16}}{\frac{5}{16} + \frac{1}{16} + \frac{1}{8}} = \frac{3}{4} \Rightarrow c = \frac{1}{3}$$

$$\text{Applying Eq. (7) with } \tau_1 = \tau_2 = 1 : x = \frac{2}{3} \Rightarrow \text{Prob}(A|B, C) = \frac{3}{5}$$

3. Monte Carlo simulation by repeating the PPM method on a set of initial realizations drawn from the prior, Prob($A|B, C$) $= 0.35$,

4. using Monte Carlo simulation on the "gradual deformation of sequential simulation" Prob$(A|B, C) = 0.27$,

The following observations can be made:

- Comparing the results of Methods 1 and 2, it is clear that the assumption of conditional independence is not valid for this case. A simple calculation shows that using $\tau_1 = 2.59$ and $\tau_2 = 1$ in Equation (7) would provide an exact approximation of the true posterior. This indicates that the single hard conditioning data should receive substantially more "weight" than assumed under conditional independence.
- While the PPM relies on the same assumption of conditional independence, the result is much closer to the true posterior probability. This result was confirmed by applying a variety of other training images, i.e. other prior distributions. Method 2 produces consistently a considerable overestimation of Prob$(A|B,C)$, the PPM is off by a few percentages only.
- It appears that the gradual deformation of sequential simulation has an implicit model of dependency between the B and C data different from the probability perturbation, and more importantly different from the actual dependence. However, different from PPM, is that in the GDM there is no τ-model that can be used to calibrate that dependence. In other words, there is no means of correcting the value of 0.27 to be closer to the true value.

The reason for the reasonably good approximation provided by the PPM can be explained by means of Equation (8). In this equation, the pre-posterior Prob$(I(\mathbf{u}_j) = 1|C)$ is a function of the data C through the parameter r_C, and, a function of an initial realization $\{i(\mathbf{u}_1),i(\mathbf{u}_2) = 1,i(\mathbf{u}_3)\}$. This initial realization depends itself on the pre-posterior Prob$(I(\mathbf{u}_j) = 1|B_j)$ with B_j depending on the random path. Hence, Equation (8) forces an explicit dependency between the Prob$(I(\mathbf{u}_j)|B_j)$ and Prob$(I(\mathbf{u}_j)|C)$ *prior* to combining both into Prob$(I(\mathbf{u}_j) = 1|B_j, C)$ using a conditional independence hypothesis (Eq. (7) with $\tau_1 = 1$, $\tau_2 = 1$). At least from this simple example, one can conclude that the *sequential* decomposition of the posterior into pre-posteriors has robustified the estimate of the true posterior under the conditional independence hypothesis.

Synthetic Example of the Single Parameter Probability Perturbation

As example application, consider the inverse problem termed "history matching." The term "history matching" in reservoir characterization and engineering describes the perturbation of a 3D reservoir model, containing petrophysical properties such as porosity and permeability, in order to match the dynamic data gathered from testing or producing the reservoir. Data consists of pressure and flow measurements in time obtained from multiple wells. In this particular case,

we will history match a facies model with known petrophysical properties per facies using water-cut data. The method can easily be extended to the case where the facies petrophysical properties are not known (see Suzuki, 2003).

Figure 3 describes in detail the synthetic case. The target model parameters **m** consist of facies indicators at each grid cell. Three facies types are present,

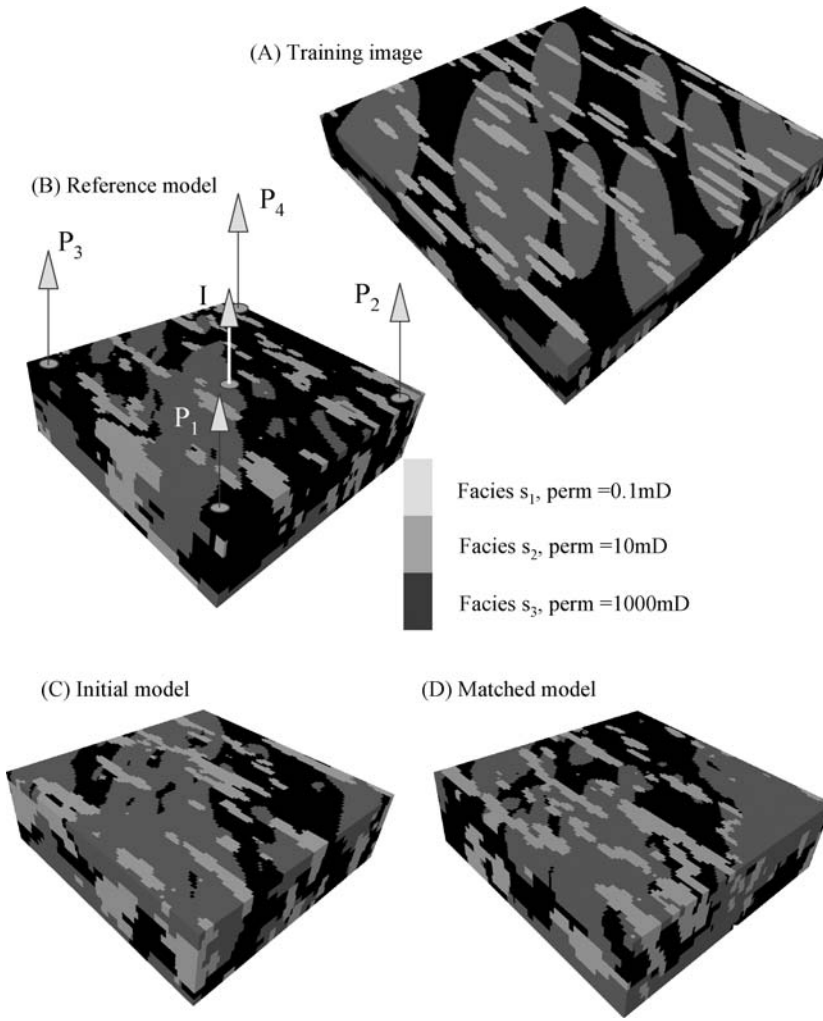

Figure 3. Demonstration example for the single parameter probability perturbation: (A) training image that defines the prior model generated by snesim, (B) the reference model with the configurations of wells, (C) an initial model that does not match the water-cut data, (D) final model that matched the water-cut.

each facies has the same constant and known porosity of 25%, but each facies has a different, constant but known permeability. The proportion of each facies is assumed to be known as well (see Hoffman and Caers, 2004, for a way to relax this assumption). A background facies s_1 has a permeability of 10 mD, an NE elliptical shaped high permeability facies s_2 of 1000 mD is eroded by NS elongated low permeability barriers s_3 of 0.1 mD. A Cartesian grid of $100 \times 100 \times 10$ cells is used. The model parameters $\mathbf{m}$ consist of the three facies indicators $I(\mathbf{u},s_1)$, $I(\mathbf{u},s_2)$, and $I(\mathbf{u},s_3)$ at each of the 100,000 grid cells.

The reservoir is initially saturated with 85% oil and 15% water, no gas is present. The water table is below the reservoir and remains so during production. The reservoir contains four producing wells in the corner of the reservoir and one injecting well in the center (see Fig. 3B). The injector pumps water in the reservoir at constant rate to drive the oil phase towards the producers. The data $\mathbf{d}$ consists of the fraction of water observed at the four producing wells up till 10 years. The forward model g is a black oil finite difference model. One single flow simulation (evaluation of the forward model) takes about 90 min.

As prior information we consider that the style of facies architecture is known, including the proportion of each facies. The multiple-point algorithm, "snesim" of Strebelle (2002) is used to simulate the facies model. This algorithm requires a 3D training image depicting the style of heterogeneity present. Figure 3A shows the training image used by snesim. Figure 3C shows an unconditional initial realization of the prior model, no hard conditioning facies data is used. The snesim algorithm was also used to generate the reference, hence we consider the prior model to be perfectly known, an assumption rarely true in reality.

To extend the binary method presented above to three (and more) categories, we re-write the key Equation (8) as follows:

$$\text{Prob}(I(\mathbf{u}_j, s_k) = 1|\mathbf{C}) = (1 - r_C) \times i_B^{(s)}(\mathbf{u}_j, s_k) + r_C P(A_j),$$
$$j = 1, \ldots, N, k = 1, .., 3$$

Otherwise the probability perturbation algorithm remains the same. Note the closure of this equation: $\sum_{k=1}^{3} \text{Prob}(I(\mathbf{u}_j, s_k) = 1|\mathbf{C}) = 1$.

The probability perturbation aims at moving the relative position of facies bodies until a history match is achieved. To start the algorithm, an initial realization is generated (see Fig. 3C). The realization is iteratively perturbed until a reasonable match is achieved. The realization obtained after 15 iterations (= random seed changes) is shown in Figure 3D, or a total of 75 flow simulations. Figure 4 shows that a reasonable match to the water-cut in the four producing wells obtained.

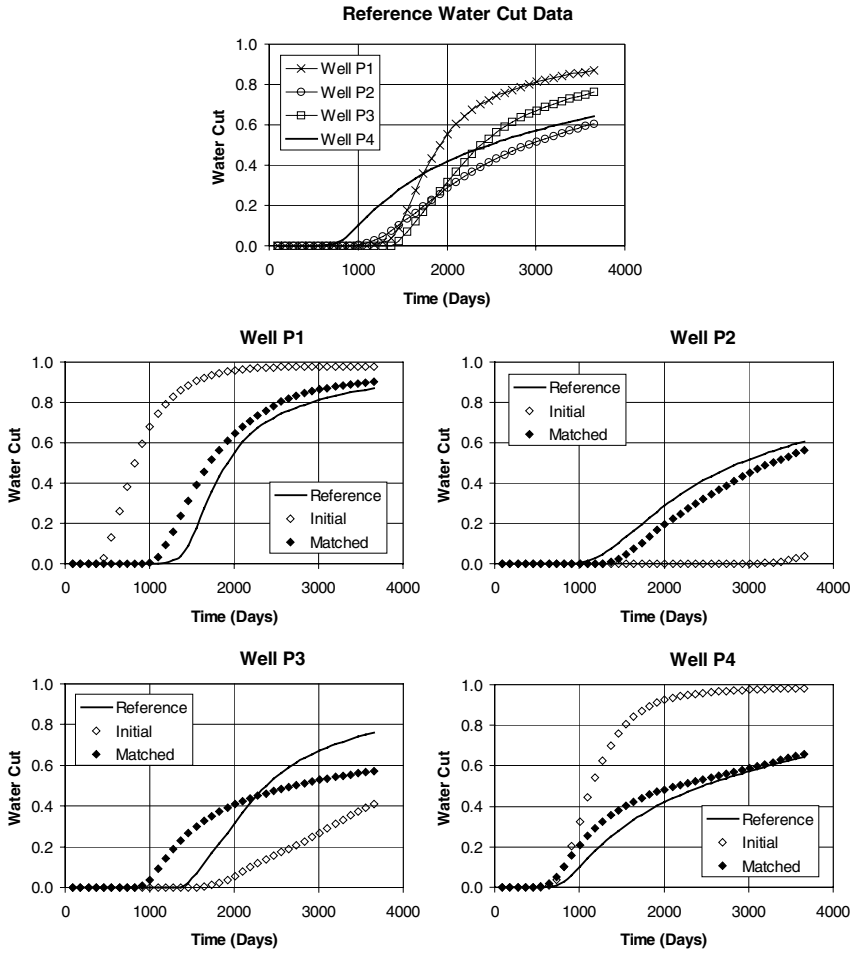

Figure 4. Single parameter probability perturbation: Water-cut curves for reference, initial, and matched model.

The regional Probability Perturbation Method

Parameterizing the pre-posterior of Equation (8) using a larger parameter set is often critical to solving large and complex inverse problems. The above—presented method does not restrict a higher order parameterization: the value of r_C can be made dependent on location $\mathbf{u}_j$

$$\text{Prob}(A_j|\mathbf{C}) = \text{Prob}(I(\mathbf{u}_j) = 1|\mathbf{C}) = (1 - r_C(\mathbf{u}_j)) \times i_B^{(s)}(\mathbf{u}_j) + r_C(\mathbf{u}_j)P(A_j),$$
$$j = 1, \ldots, N \tag{10}$$

The use of (8) in the probability perturbation method now requires a multidimensional optimization on all $r_C(\mathbf{u}_j)$, $j = 1, \ldots, N$. The elegance by which Equation (8) reduces an N-dimensional to a one-dimensional search has therefore been lost.

To avoid a potentially difficult full multi-dimensional search for the best $r_C(\mathbf{u}_j)$, $j = 1, \ldots, N$, we will rely on a region-wise parameterization of these parameters. Consider L regions in the domain of study, each region R_ℓ, $\ell = 1, ..., L$ consists of a set of N_ℓ grid node locations,

$$R_\ell = \left\{\mathbf{u}_i^{(\ell)}, \mathbf{u}_j^{(\ell)}, \ldots\right\}$$

The grid nodes in one single region need not be connected, but each grid node must belong to only one region. Which nodes belong to which region is a problem–specific question. The number of regions L however is likely to be considerably less than the number of grid nodes N. The pre-posterior of Equation (10) is rewritten using a region-wise parameterization as follows:

$$\text{Prob}\left(A_j^{(\ell)}|\mathbf{C}\right) = \text{Prob}\left(I\left(\mathbf{u}_j^{(\ell)}\right) = 1|\mathbf{C}\right) = \left(1 - r_C^{(\ell)}\right) \times i_B^{(s)}\left(\mathbf{u}_j^{(\ell)}\right) + r_C^{(\ell)} P\left(A_j^{(\ell)}\right),$$
$$j = 1, \ldots, N$$

where the parameter $r_C^{(\ell)}$is the same for all grid nodes $\mathbf{u}_j^{(\ell)}$ of region R_ℓ. Note also that the prior $P(A_j^{(\ell)})$ need not be the same for all regions. An efficient strategy for jointly optimizing on all L $r_C^{(\ell)}$ parameters is discussed in Hoffman and Caers (2003).

To demonstrate the benefit of the higher dimensional parameterization, we revisit the synthetic reservoir example shown above. Four rectangular regions are created with the injector at the center (see Fig. 5A). The regionwise perturbation allows the facies bodies to be perturbed by different amounts in each region, however, without creating any discontinuities at the regions borders (see Hoffman and Caers, 2004, for a detailed explanation). Since flow between each producer and injector takes place mostly in the quadrant occupied by the producer, the facies distribution in each quadrant is mostly consequential to the flow data in that quadrant, hence the efficiency gain expected over the single parameter method. Indeed, the number of iteration required to reach the same accuracy of the match (see Fig. 6) as in the one-parameter case equals 3 (21 flow simulations). A single matched model is shown in Figure 5C, note that despite the existence of regions, the final model displays the same style of continuity as the reference.

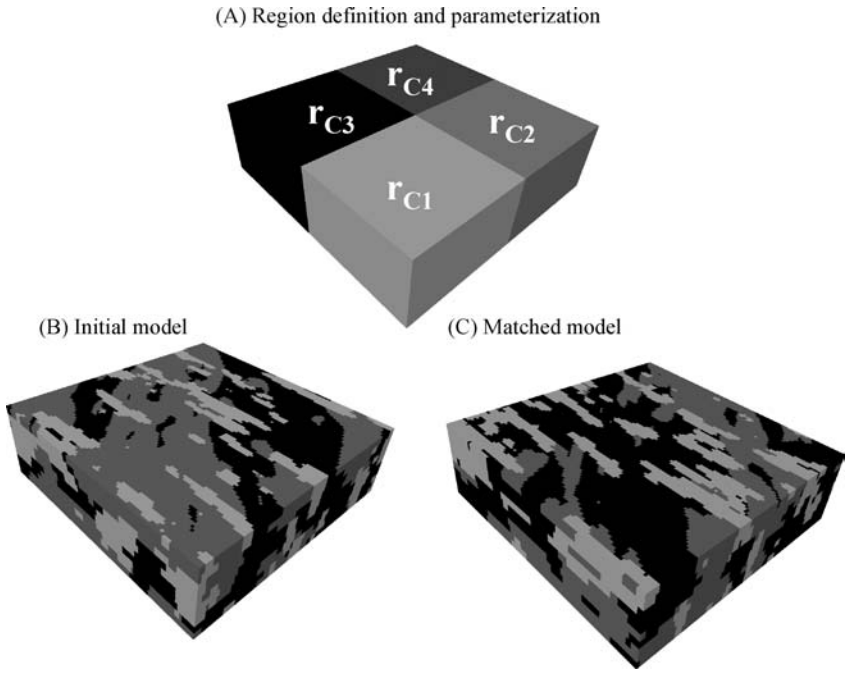

Figure 5. Demonstration example for the regional probability perturbation: (A) definition of the regions, to each region a perturbation parameter r_C is assigned, (B) the same initial model as in the single parameter case, and (C) final model that matched the water-cut.

DISCUSSION AND CONCLUSIONS

The probability perturbation has a goal no different from any other Bayesian inverse method: drawing realizations that honor the data and honor prior statistics. In the traditional Bayesian inversion methods, this is done by defining a prior and a likelihood. The probability perturbation method relies on the principle of sequential simulation to sample the posterior distribution and relies on pre-posteriors instead of likelihoods.

Nevertheless, the probability perturbation method shares two important properties of other inverse algorithms:

1. Inverse solutions generated using the probability perturbation method honor the prior statistics. Whatever the value of r_C or random seed s, all model realizations are drawn using a sequential sampling method that depends on Prob($A_j|\mathbf{B}$) (depending on the prior statistics). The probability perturbation is essentially a search method for finding those prior model realizations that honor the data $\mathbf{d}_2$.

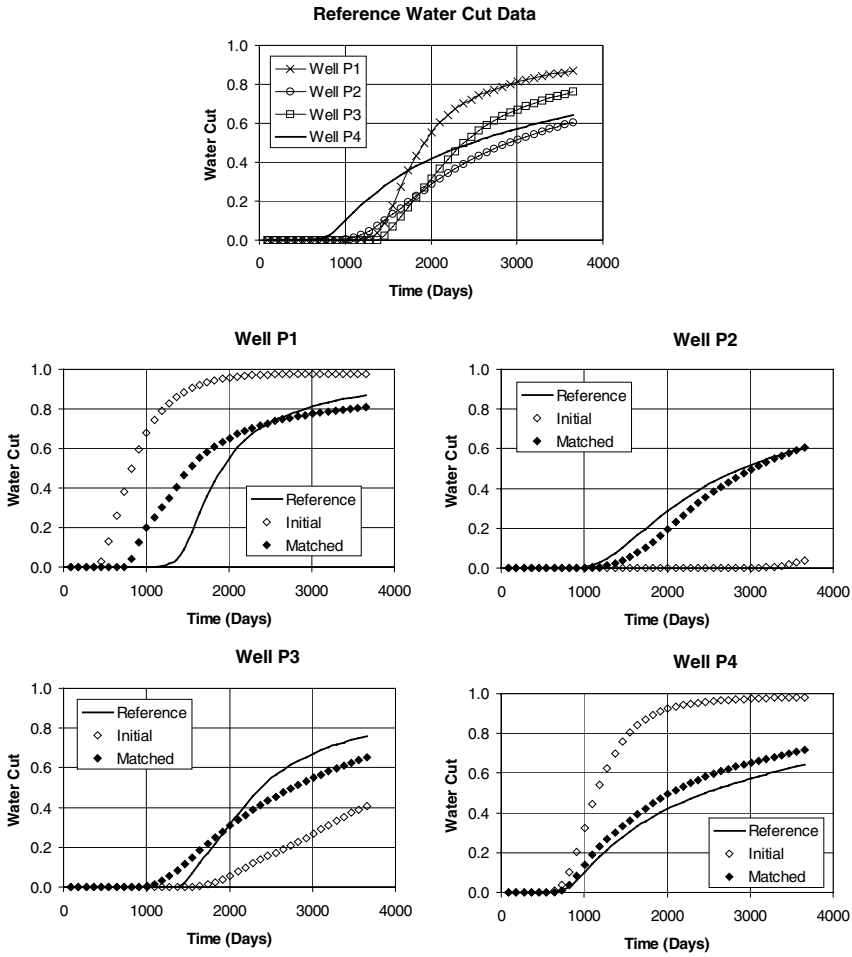

Figure 6. Regional probability perturbation: Water-cut curves for reference, initial, and matched model.

2. The probability perturbation method can cover the entire space of the prior model realizations. At each step of the outer loop, the random seed is changed, hence, if infinite computing time were available, the method would cover the entire space of possible realizations, just like a rejection sampler would.

These two important properties do not reflect how efficient the probability perturbation method works. The parameterization of the pre-posterior using a single parameter is most likely inadequate for finding solutions of complex problems

in a reasonable amount of time. To make the method feasible for solving large and complex problems, one most likely requires a larger-dimensional parameterization, as done in the regional probability perturbation method. This paper reflects the theory behind the probability perturbation, for large-scale applications to actual reservoirs, the reader is referred to some papers in the reference list.

REFERENCES

Besag, J., and Green, P. J., 1993, Spatial statistics and Bayesian computation: J. R. Stat. Soc. B, v. 55, p. 3–23.

Caers, J., 2003, History matching under a training image-based geological model constraint: Soc. Petrol. Eng. J., SPE #74716, p. 218–226

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: The geostatistical software library: Oxford University Press, New York, 369 p.

Hegstad, B. K., and Omre, H., 1996, Uncertainty assessment in history matching and forecasting, *in* Baafi, E. Y., and Schofield, N. A., eds., Geostatistics Wollongong, Kluwer, Dordrecht, v. 1, p. 585–596.

Hoffman, B. T., and Caers, J., 2003, Geostatistical history matching using the regional probability perturbation method, *in* Proceedings of the SPE Annual Conference and Technical Exhibition, Soc. Petrol. Eng., Paper no. 84409, Denver, CO, October 5–8, 16 p.

Hoffman, B. T., and Caers, J., 2004, History matching with the regional probability perturbation method—Applications to a North Sea reservoir, *in* Proceedings of the ECMOR IX, Cannes, August 29 – September 2, 2004.

Hu, L.-Y., Blanc, G., and Noetinger, B., 2001, Gradual deformation and iterative calibration of sequential simulations: Math. Geol., v. 33, p. 475–489.

Journel, A. G., 1993, Geostatistics: Roadblocks and challenges, *in* Soares, A., ed., Geostatistic-Troia: Kluwer Academic, Dordrecht, v. 1, p. 213–224.

Journel, A. G., 2002, Combining knowledge from diverse data sources: An alternative to traditional data independence hypothesis: Math. Geol., v. 34, p. 573–596.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, Equation of state calculation by fast computing machines: J. Chem. Phys., v. 21, p. 1087–1092.

Moosegard, K., and Tarantola, A., 1995, Monte Carlo sampling of solutions to inverse problems: J. Geophys. Res. B, v. 100, p. 12431–12447.

Neal, R. M., 1993, Probabilistic inference using Markov chain Monte Carlo methods: Technical report CRG-TR-93-1, Department of Computer Science, University of Toronto.

Omre, H., and Tjelmeland, H., 1996, Petroleum geostatistics, *in* Baafi, E.Y., and Schofield, N. A. eds., *in* Geostatistics Wollongong: '96 Kluwer, Dordrecht, v. 1, p. 41–52.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point geostatistics: Math. Geol., v. 34, p. 1–22.

Suzuki, S., 2003, Determining petrophysical properties of facies using a hierarchical history matching method: MSc Thesis, Stanford University, Stanford, CA.

Tarantola, A., 1987, Inverse problem theory: Elsevier Science, Amsterdam, 613 p.