Technical University of Denmark

DTU

Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling

Hansen, Thomas Mejer; Cordua, Knud Skou; Mosegaard, Klaus

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Inverse problems with non-trivial priors: Efficient solution through Sequential Gibbs Sampling

Thomas Mejer Hansen · **Klaus Mosegaard** · **Knud Skou Cordua**

Abstract Markov chain Monte Carlo methods such as the Gibbs sampler and the Metropolis algorithm can be used to sample solutions to non-linear inverse problems. In principle these methods allow incorporation of prior information of arbitrary complexity. If an analytical closed form description of the prior is available, which is the case when the prior can be described by a multidimensional Gaussian distribution, such prior information can easily be considered. In reality, prior information is often more complex than can be described by the Gaussian model, and no closed form expression of the prior can be given. We propose an algorithm, called sequential Gibbs sampling, allowing the Metropolis algorithm to efficiently incorporate complex priors into the solution of an inverse problem, also for the case where no closed form description of the prior exists. First we lay out the theoretical background for applying the sequential Gibbs sampler and illustrate how it works. Through two case studies we demonstrate the application of the method to a linear image restoration problem and to a non linear cross borehole inversion problem. We demonstrate how prior information can reduce the complexity of an inverse problem, and that a prior with little information leads to a hard inverse problem, practically unsolvable except when the number of model parameters is very small. Considering more complex and realistic prior information thus not only makes realizations from the posterior look more realistic, it can also reduce the computation time for the inversion dramatically. The method works for any statistical model for which sequential simulation can be used to generate realizations. This applies to most algorithms developed in the geostatistical community.

Center for Energy Resources Engineering, DTU Informatics, Technical University of Denmark, Lyngby, Denmark
Tel.: +45-4525 3351
Fax: +45-4588 2673
E-mail: tmeha@imm.dtu.dk

1 Introduction

Consider a forward problem

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

where a function f relates a subsurface model $\mathbf{m}$ to observational data $\mathbf{d}$. Inverse problem theory deals with the problem of inferring properties of $\mathbf{m}$ from a specific data set $\mathbf{d}$, using equation (1) and possibly some prior information on $\mathbf{m}$.

Tarantola and Valette [47] formulated a probabilistic approach to solving inverse problems where all available states of information is described by probability density functions (pdfs). The solution to the inverse problem is the probability distribution obtained combining all the known states of information. In a typical inverse problem the states of information can be described by the *prior* pdf and the *likelihood* function. The prior pdf, $\rho_{\mathrm{M}}(\mathbf{m})$, describes the data-independent prior knowledge of the model parameters. The likelihood function, $L(\mathbf{m})$, is a probabilistic measure of how well the data associated to a given model match a given model of data uncertainty. The solution to such an inverse problem is then the posterior pdf, which is proportional to the product of the prior and the likelihood:

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k \, \rho_{\mathrm{M}}(\mathbf{m}) \, L(\mathbf{m}) \,, \tag{2}$$

where k is a normalization factor. The posterior pdf describes all models consistent with both prior information and data. For some inverse problems the complete posterior pdf can be obtained analytically (such as for example linear inverse Gaussian problems). For other inverse problems the only way to characterize the posterior pdf is by sampling it [33]. A sample of the posterior is a set of models (realizations of the posterior) where each model occurs with a frequency

proportional to its posterior likelihood. The frequency by which a particular feature of a model appears in the posterior sample is a measure of the probability of that feature existing, according to prior information and data. Sampling the posterior thus allows relatively sophisticated analysis of the posterior pdf [32]. The movie strategy, advocated by Tarantola [46], propose to visualize samples from the prior and the posterior as movies. The 'prior movie' will make it apparent what prior choices have been made. The difference between the prior and the posterior movie will highlight the effect of using data.

In principle prior information of arbitrary complexity can be included in the solution of the probabilistically framed inverse problem of Eq. 2. In practice though, the application of inverse problems with for example geologically realistic prior information have been limited, Tarantola [46] page 52. There may be several reasons for this.

One reason is that the use of prior information itself has been the source of a heated debate [22,34,41]. Some authors suggest that the use of prior information may bias the solution of an inverse problem in an unwanted way. Others adopt the probabilistic viewpoint of Tarantola and Valette [47], but propose to keep prior information to a minimum in order to solve the problem without biasing the inversion result, see e.g. Buland and Omre [5] and Khan and Mosegaard [27]. This corresponds to the idea of using a 'noninformative' prior model [3,42]. Jaynes and Bretthorst [23] advocate using a prior based on maximum entropy, which contains the least information consistent with prior constraints.

In this paper we will stress and exemplify one important aspect of this discussion: When prior information $\rho_M(\mathbf{m})$ is consistent with information contained in the data, that is, when regions of significant probability/likelihood of $\rho_M(\mathbf{m})$ and $L(\mathbf{m})$ clearly overlap, the hardness of the inverse problem tends to decrease. On the other hand, the use of inconsistent prior information may render an otherwise easy inverse problem hard, and in practice unsolvable.

Another reason for the use of relatively simple prior models, is that, until recently, algorithms have not been available to efficiently and accurately quantify complex prior information, such as for example geologically realistical patterns, in a probabilistic form. An exception is the case where both the prior and the model of data noise can be described by Gaussian statistics, and the inverse problem is linear (or linearizable). In this case the posterior pdf can be described by a Gaussian pdf fully characterized by a mean model and a covariance. A sample from the posterior pdf (and the a prior pdf as well), can be generated using for example Cholesky decomposition, or sequential simulation, Hansen et al. [17]. Thus for the linear inverse Gaussian case, the movie strategy is today practically feasible. See for example Buland and Omre [5] for an example of probabilistic based linear Gaussian inversion.

The linear and Gaussian assumptions are convenient as they lead to computationally feasible inversion algorithms. In reality though, most geophysical inverse problems are non-linear, and the prior choice of Gaussian statistics to describe both the noise model and the prior model distribution is extremely limited.

Sampling methods such as the Metropolis Algorithm [20, 31] can be use to sample any probability density function, hence also the posterior pdf. In its original form the Metropolis algorithm can be implemented in the following way in order to the sample the posterior:

The Metropolis algorithm

If we have a way of evaluating the values of both $L(\mathbf{m})$ and $\rho_M(\mathbf{m})$ at any point $\mathbf{m}$ in model space $\mathcal{M}$, and an algorithm A (a proposal generator) that is able to sample $\mathcal{M}$ at random, the following algorithm will sample the posterior (Eq. 2):
- Starting in the current model $\mathbf{m}_c$, perform one step with the uniform sampler A.
- Accept the new point $\mathbf{m}_t$ only with probability
$$P_{accept} = \min\left(1, (L(\mathbf{m}_t)\rho_M(\mathbf{m}_t))/(L(\mathbf{m}_c)\rho_M(\mathbf{m}_c))\right). \quad (3)$$
- If $\mathbf{m}_t$ is rejected, re-use $\mathbf{m}_c$ in the next step.
- In case $\mathbf{m}_t$ is accepted, let $\mathbf{m}_c = \mathbf{m}_t$ in the next step.

This classical Metropolis sampler require that both $(L(\mathbf{m})$ and $\rho_M(\mathbf{m})$ can be evaluated. While this algorithm in principle is easy to implement it may not be trivial to compute $\rho_M(\mathbf{m})$ from complex prior models. The Metropolis algorithm may also be computationally demanding when the prior model is far from a uniform.

Mosegaard and Tarantola [37] propose a Markov chain algorithm based on the Metropolis Algorithm [20,31], that we will refer to as the *extended Metropolis algorithm*, that can be implemented in the following way:

Extended Metropolis algorithm

If we have a way of evaluating the values of $L(\mathbf{m})$ and an algorithm B (a prior generator) that is able to sample $\rho_M(\mathbf{m})$ directly (without necessarily evaluating $\rho_M(\mathbf{m})$ anywhere), the following algorithm will sample the posterior (Eq. 2):
- Starting in the current model $\mathbf{m}_c$, perform one step with the prior sampler B.
- Accept the new point $\mathbf{m}_t$ only with probability
$$P_{accept} = \min(1, L(\mathbf{m}_t)/L(\mathbf{m}_c)). \quad (4)$$
- If $\mathbf{m}_t$ is rejected here, re-use $\mathbf{m}_c$ in the next step.
- In case $\mathbf{m}_t$ is accepted, let $\mathbf{m}_c = \mathbf{m}_t$ in the next step.

The extended Metropolis sampler allows sampling the posterior pdf for non-linear inverse problems in presence of an arbitrarily complex prior model considering any noise model for the data.

Algorithms based on the classical Metropolis sampler are the easiest to implement because it is simple to sample

$\mathcal{M}$ uniformly, and because it is simple to evaluate $\rho_M(\mathbf{m})$ when a formula for ρ_M is available. In contrast to this, algorithms based on the extended Metropolis sampler may be more difficult to implement, because knowledge of $\rho_M(\mathbf{m})$ must be built directly into the sampler B [36]. However, because knowledge of $\rho_M(\mathbf{m})$ is built directly into the sampler B, algorithms based on the extended Metropolis sampler can be dramatically more computationally efficient than algorithms based on the classical Metropolis algorithm [37]

The extended Metropolis algorithm is, in principle, simple to use and its versatility appealing. Yet it has mostly been applied to sample the posterior of non-linear inverse problem with relatively simple prior models. One reason could be that until recently few methods have existed allowing sampling of complex prior models. Khan and Mosegaard [27] and Voss et al. [49] use the extended Metropolis algorithm using uniform Gaussian priors, with no spatial correlation between model parameters. Bosch [2] consider a Gaussian prior model with spatially correlated model parameters.

More complex priors based on Markov random fields have been developed for Bayesian image analysis, see e.g. Besag [1] and Tjelmeland and Besag [48]. These algorithms tend to be iterative and relatively computationally demanding.

One of the most promising ways of using the extended Metropolis Algorithms will undoubtedly be to incorporate geostatistical information into the solution of geophysical inverse problems.

Geostatistics is a an application of random function theory to characterize natural phenomena, Journel and Huijbregts [26] page 1. For several decades the geostatistical community has been developing algorithms and methods that are able to model increasingly complex geological features. Most of these methods are based on random function theory [14, 16, 25, 8, 43–45, 38], and most are available through the SGeMS software package, Remy et al. [40]. These methods can all in principle be used to quantify prior beliefs of the spatial variability of the subsurface in a statistical sound manner in form of probability densities. Realizations of the various models can be generated using sequential simulation [14].

Hansen et al. [19] suggest a simple algorithm that allows using any random function model, that can be sampled using sequential simulation, to quantify the prior information. We shall show that it can be used with the extended Metropolis algorithm.

Here we will develop the theory behind the algorithm and show that it can be seen as an application of a combination of the Gibbs sampler and sequential simulation. Hence we propose to refer to the algorithm as 'sequential Gibbs sampling'.

We will lay out the theoretical background for applying the sequential Gibbs sampler, proving that it will sample the random function model intended, and illustrate how it

works, as part of the extended Metropolis sampler. Through two case studies we demonstrate the application of the method to a linear image restoration problem and to a non linear cross borehole inversion problem.

2 Quantifying prior information using geostatistics

Geostatistical random function models can be divided into two groups based on 2-point statistics and multiple-point statistics. Traditional geostatistical random function models rely on 2-point statistics, where spatial variability is only described between pairs of two data locations, typically quantified by a covariance model. The simplest 2-point based geostatistical model is the Gaussian model which can be completely described by a mean and a covariance model, with an implicit assumption of a Gaussian distribution of model parameters. Realizations of such a Gaussian model can be generated using sequential Gaussian simulation [14]. Realizations from random function models based on a 2-point statistical model defined by a mean and a covariance, but with an arbitrary distribution of model parameters, can be obtained using direct sequential simulation [8,25,43].

While somewhat complex structures can be quantified using the 2-point based simulation algorithms, geological realistically features such as channels cannot. To model more complex features higher order statistical moments must be considered [24]. Such higher order statistical models are typically referred to as models based on multiple-point statistics. Guardiano and Srivastava [16] propose an algorithm that can simulate spatial features consistent with a higher order statistical model inferred from a training image. The algorithm was however not computationally feasible, and it was not until Strebelle [44,45] proposed the single normal equation simulation algorithm (SNESIM) that the method became practical to use. Lately Mustpapha and Dimitrakopoulos [38] proposed a multiple-point based algorithm allowing reproducing a number of higher order cumulants as observed from data, or from a training image.

All these methods can in principle be used to quantify prior beliefs of the spatial variability of the subsurface in a statistical sound manner in form of probability densities, and realizations of the various models can be generated using sequential simulation.

2.1 Sequential simulation

$$(5)$$

Consider an image of the subsurface, consisting of N pixels (voxels), each characterized by a physical or geological parameter m_i. A joint probability density $f_M(m_1 \ldots m_N)$ defines a random field describing the correlations between

parameters. A realization from the random field can be simulated using a technique knows as 'sequential simulation' as follows: In N steps, visit each parameter (pixel) sequentially. In step i, visit parameter m_i and generate a realization of m_i from the conditional probability density function

$$f_M(m_i|m_1 \ldots m_{i-1}), \tag{6}$$

That the above procedure will actually generate a realization from $f_M(m_1 \ldots m_N)$ follows from the identity $f(s|t)f(t) = f(s,t)$, which (in the general multivariate case) yields

$$\begin{aligned} f_M(m_{n+1} \ldots m_N|m_1 \ldots m_n) &= f_M(m_{n+1}|m_1 \ldots m_n) \\ &\quad f_M(m_{n+2}|m_1 \ldots m_{n+1}) \\ &\quad \vdots \\ &\quad f_M(m_N|m_1 \ldots m_{N-1}). \end{aligned} \tag{7}$$

When all locations have been visited one realization is generated. Thus to apply sequential simulation one must a) build a local conditional pdf (conditional to the previously simulated data), and b) draw a realization of this local pdf.

More detailed descriptions of the theory and application of sequential simulation can be found in e.g. Gómez-Hernández and Journel [14] and Goovaerts, chapter 8.2 [15].

Considerable effort have been made in the geostatistical community to efficiently compute conditional probability density functions, as in Eq. 6, based on 2-point [14, 39, 43] and multiple-point stochastic models [40, 45, 50].

2.2 Gibbs sampling

Consider a known realization $\mathbf{m}$ of the random field described by the probability distribution $f_M(m_1 \ldots m_N)$. If we randomly select a model parameter, $m_i = m(\mathbf{u}_i)$, compute the local conditional pdf

$$\rho_M(m_i|m_1, m_2, \ldots, m_{i-1}, m_{i+1}, \ldots, m_N). \tag{8}$$

and draw a value from it, we get a new realization of the random field defined by $f_M(m_1 \ldots m_N)$. If this is repeated iteratively, it will be an application of the Gibbs sampler [13], and thus allow sampling from $f_M(m_1 \ldots m_N)$. Note that the Gibbs sampler only requires that a sample from ρ_M can be generated. The full conditional pdf of Eq. 8 need not be computed.

2.3 Sequential Gibbs sampling

The cost of using the Gibbs sampler is that one must be able to generate a realization of the local conditional pdf, Eq. 8, which can be done very effectively using methods developed for sequential simulation, eqn. 6-7. We therefore suggest to combine sequential simulation and Gibbs sampling, and refer to this combination as Sequential Gibbs Sampling. Using

the sequential Gibbs sampler we will be able to sample realizations of the probability distribution $f_M(m_1 \ldots m_N)$.

The sequential Gibbs sampler can be used as a prior sampler for the extended Metropolis algorithm (Mosegaard and Tarantola, 1995). However, in order to control the computational efficiency the algorithm, some flexibility in the degree of perturbation (the 'step-length') is needed. A step length leading to a frequency of acceptance rates of about 25% to 50% is considered to provide a computational efficient Metropolis sampler [11, 33]. We suggest considering not just one model parameter at each step of the Gibbs sampler, but a subset U of all model parameters. Assuming that $\mathbf{m}_{i \in U}$ contains the selected parameters to be perturbed, and that $\mathbf{m}_{i \notin U}$ contains the remaining parameters, we now draw a realization of $\mathbf{m}_{i \in U}$ from $\rho_M(\mathbf{m}_{i \in U}|\mathbf{m}_{i \notin U})$. In Appendix A it is demonstrated that this procedure, if applied iteratively, will sample the distribution $\rho_M(\mathbf{m})$.

To draw a realization from $\rho_M(\mathbf{m}_{i \in U}|\mathbf{m}_{i \notin U})$ we could compute $\rho_M(\mathbf{m}_{i \in U}|\mathbf{m}_{i \notin U})$ of $\mathbf{m}_{i \in U}$ explicitly. This is, however, extremely inefficient, and fortunately not necessary, as we only need to be able to generate a realization of $\rho_M(\mathbf{m}_{i \in U}|\mathbf{m}_{i \notin U})$ not the full conditional pdf itself. Instead, we can use the sequential simulation approach of Eq. 6, which involves computing only the conditional probability density function for each model parameter in U in random order. In this way we have designed an efficient Gibbs sampler that is able to incorporate complex prior information and, at the same time allows us to control the 'step length', and hence the efficiency, of the sampling.

2.4 An algorithm for sequential Gibbs sampling

Implementing the sequential Gibbs sampler amounts to implementing a Gibbs sampler which, in each iteration, calculates a realization of the conditional probability density function associated to a specific subset of model parameters using sequential simulation:

1. Select a subset of the model parameters, U, and regard these as unknowns. The rest of the model parameters are considered known (and fixed).
2. Perform sequential simulation of the unknown parameters conditioned to the known parameters. This generates a new model, which is also a realization of the prior model. This step is identical to drawing a value from the conditional probability density function in Eq. 6.
3. Use the new model as the starting model and go to (i).

As already mentioned, the number of model parameters (size of U) to be resimulated (in step 1. in the sequence above) can be used to control the 'step length' of the sequential Gibbs sampler, which is essential to computational efficiency if the method is used as part of a Metropolis sampling algorithm. Resimulating only a single model parameter results

in a model $\mathbf{m}_{i+1}$ that is highly related to the original model $\mathbf{m}_i$. On the other hand, resimulating all model parameters leads to an $\mathbf{m}_{i+1}$ that is statistically independent of m_i.

Figure 2 illustrates 3 examples of using a random walk to generate realizations of a prior model defined by the spatial statistics of the training images in Figure 1. We make use of the SNESIM algorithm to compute a realization of the conditional distribution Eq. (8) [45].

Figures 2a and 2b make use of the channel based training image of Figure 1a, using a resimulation area, U, of 12x12 and 4x4 cells respectively. Note that U does not need to take any specific shape. Any subset of model parameters can be considered. Figures 2c is based on the prior model associated with to the training image in Figure 1b. Progress of the sequential Gibbs sampling is from left to right, where the initial realization is to the left. The area in gray colors indicate the model parameters that is to be resimulated. In the following realization (to the right) the data within the grey area have been resimulated and new area is selected for conditional resimulation. The final realization after 6 iterations of the sequential Gibbs sampler is the model to the right.

The training images of Figure 2 are but two examples of a categorical 2-facies training image. For numerous examples of algorithms and associated training images see Remy et al. [40].

A crucial step for applying simulation based on training images, and hence the use of the sequential Gibbs sampling, is the existence of realistic training images. For geophysical inverse problems prior information is often available as a conceptual geological model. See for example Zhang [52] for an example on how to quantify such a geological conceptual model by training images. Using a different approach Zhang et al. [53] demonstrate how a 3D training image of sandstone can be computed from 2D sections.

[Fig. 1 about here.]

[Fig. 2 about here.]

Related work

As mentioned earlier the 'resampling' algorithm presented above was originally proposed by Hansen et al. [19], and subsequently an almost identical method was proposed by Mariethoz et al. [30]. Neither of these works present a theoretical background for using the method, and provide no proof that the resulting algorithm samples an equilibrium distribution, nor that such an equilibrium distribution would in fact be the requested prior model.

Mariethoz et al. [30] essentially describes an optimization application of the resampling technique used together with a version of the Metropolis sampler, where only models that increase the likelihood are accepted.

Fu and Gómez-Hernández [9,10] propose a 'blocking' Markov chain Monte Carlo method, related to multi-Gaussian conditional simulation. In each step of the Metropolis algorithm a 'block' of coherent model parameters is selected. Model parameters at the edge of the chosen block, is retained as conditioning data, and the the rest of the data in the block is simulated conditional to the edge data.

Without describing it in detail, Bosch [2] also seem to have made use of a technique similar to re-simulation of one data point at a time as given in Eq. 6 for a Gaussian probability distribution. Irving and Singha [21] make use of resimulation to sample the prior using the sequential indicator simulation algorithm.

Here we have provided the formal proof that such an approach is valid for any probability distribution that can be sampled using sequential simulation. We shall further investigate how the use (and lack of) prior information affect the computational complexity of inverse problems.

3 Application examples

Sequential Gibbs sampling has the potential to allow the Metropolis algorithm to sample solutions to inverse problems with relatively complex prior models. We will demonstrate the applicability of the method using two classical inverse problems: image de-blurring and tomography.

3.1 Image de-blurring

Consider the 41x41 pixel gray-scale image of Figure 3 as a reference image. It has been generated using single normal equation simulation (Strebelle, 2002) using the training image in Figure 1a. A Gaussian kernel with a horizontal range of 15 grid cells and a vertical range of 6 grid cells, see Figure 3b, is used as a smoothing kernel to obtain a smoothed version of the reference gray scale image in Figure 3a, as seen in Figure 3c. Then uncorrelated Gaussian noise with a standard deviation of 0.045 (reflecting a signal to noise ratio of 0.1) is added to the data, Figure 3d. We now consider the 5x5 pixels indicated by circles in Figure 3d as observed data.

[Fig. 3 about here.]

Reconstructing the 41x41 gray scale image, Figure 3a, from the 5x5 pixel data, can now be seen as a simple linear inverse problem, where the forward problem is described by

$$\mathbf{d} = \mathbf{Gm} \qquad (9)$$

where $\mathbf{G}$ describes the smoothing kernel. A model of the noise can be described by a data covariance, $\mathbf{C}_D$, which is the matrix with a constant value of 0.045 in the diagonal.

We consider three choices of prior models of the spatial distribution of the model parameters, $\rho_{\mathrm{M}}^{\mathrm{TI}}$, $\rho_{\mathrm{M}}^{uncorr}$, and ρ_{M}^{corr} defined as

– $\rho_{\mathrm{M}}^{\mathrm{TI}}$

   The statistical model described by the reference training image, Figure 1a. This may be the ideal choice of prior model to use, as we know that the reference image we try to reconstruct, is a realization of $\rho_{\mathrm{M}}^{\mathrm{TI}}$.

– $\rho_{\mathrm{M}}^{uncorr} \rightarrow N(\mathbf{m}_0, \mathbf{C}_{\mathbf{m}^{uncorr}})$

   A Gaussian prior model with no spatial correlations. From the reference model we compute the mean value for all pixels as $\mathbf{m}_0$=0.28. As no spatial correlation is assumed the model covariance matrix can be described by the identity matrix times the variance of the original pixel data, $\mathbf{C}_{\mathbf{m}^{uncorr}} = \mathbf{I}\,0.45^2$.

– $\rho_{\mathrm{M}}^{corr} \rightarrow N(\mathbf{m}_0, \mathbf{C}_{\mathbf{m}^{corr}})$

   From the reference image of Figure 3a, we estimate an apparent covariance model, $\mathbf{C}_{\mathbf{m}^{corr}}$, as an anisotropic spherical covariance model with a horizontal range of 18 pixels and a vertical range of 6 pixels. The prior mean is as for $\rho_{\mathrm{M}}^{uncorr}$, $\mathbf{m}_0$=0.28.

All three considered prior models are consistent with the spatial statistics of the reference image up to a certain order. $\rho_{\mathrm{M}}^{uncorr}$ is the least informed prior model, reflecting only the lower order mean and variance, while ρ_{M}^{corr} also reflect the correct covariance. $\rho_{\mathrm{M}}^{\mathrm{TI}}$ is the most informed prior model, reflecting higher order spatial statistical features, such as the channels observed in the training image.

Note that using the training image based prior model implies a prior assumption that the pixel values can only attain the values 0 and 1. Assuming Gaussian type prior models implies a prior assumption that the pixel values are continuous values and normally distributed according to the prior covariance model.

3.1.1 Linear least squares inversion

For Gaussian type prior models, the inverse problem can be directly solved using least squares inversion. As the forward problem is linear and both data and model covariance is given by Gaussian statistics, the full posterior distribution is a Gaussian probability density function, fully characterised by the posterior mean and covariance, N($\widetilde{\mathbf{C}}_{\mathrm{M}}$) (from Tarantola, 2005):

$$\widetilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_{\mathrm{M}}\mathbf{G}^t\,(\mathbf{G}\,\mathbf{C}_{\mathrm{M}}\,\mathbf{G}' + \mathbf{C}_{\mathrm{D}})^{-1}\,(\mathbf{d}_0 - \mathbf{G}\,\mathbf{m}_0) \quad (10)$$

$$\widetilde{\mathbf{C}}_{\mathrm{M}} = \mathbf{C}_{\mathrm{M}} - \mathbf{C}_{\mathrm{M}}\mathbf{G}^t\,(\mathbf{G}\,\mathbf{C}_{\mathrm{M}}\,\mathbf{G}' + \mathbf{C}_{\mathrm{D}})^{-1}\,\mathbf{G}\,\mathbf{C}_{\mathrm{M}} \quad (11)$$

To obtain actual realizations of the posterior we use Cholesky decomposition of $\widetilde{\mathbf{C}}_{\mathrm{M}}$. In this way four independent realizations have been generated, using the prior models of $\rho_{\mathrm{M}}^{uncorr}$ and ρ_{M}^{corr} respectively as seen on Figure 4a and 4b.

Perhaps not surprisingly the prior assumption of no spatial correlation results in posterior realizations with very little spatial structure, and certainly no channel like structures. Assuming the spatially correlated covariance model, which is in fact consistent with the 2-point statistical properties of the reference training image, some channel like structures appear, but the apparent look of the realizations is far from the channel like structures of the reference image and the training image. This is probably related to the fact that Gaussian simulation is a maximum entropy algorithm, that will lead to maximum disorder for higher order moments. Therefore Gaussian simulation should not be expected to reproduce spatial features that can only be described by higher order statistical moments [24].

3.1.2 Non-Gaussian prior information

Considering now the prior model defined by the training image, $\rho_{\mathrm{M}}^{\mathrm{TI}}$, least squares inversion cannot be used, as the model covariance cannot be described by Gaussian statistics. Instead we use the extended Metropolis algorithm to sample the posterior probability distribution. The sequential Gibbs sampler, as described previously, is used to sample the prior model.

As a starting model we use the smoothed noisy gray scale image of Figure 3d where all pixel values below .28 is set to 0, and all pixel values above .28 is set to 1. This provides a starting model consistent with the prior choice of only ones and zeros, but where the spatial distribution is not a realization consistent with the higher order statistics from the training image, see Figure 4c, first model.

The extended Metropolis algorithm is run for 10000 iterations producing approximately 100 independent realizations of the posterior probability distribution. Three of these is shown in Figure 4c. Note how all these realizations contain channel like structures, with channel thicknesses as expected from the prior model. It is evident comparing the posterior realizations in Figure 4a-c, that realizations based on the training image are superior to the realizations based on Gaussian statistics in terms of reproducing the structures of the reference image, Figure 3a.

3.1.3 Posterior probability of a channel

An alternative to showing realizations of the posterior is to show for example the average of all possible models, which in the least squares case coincide with the model of maximum posterior probability, $\widetilde{\mathbf{m}}$. For the Gaussian based prior models Figures 5a-b show this model with maximum posterior probability. Compared to the reference image of Figure 3a, it is clear that using the spatially correlated prior covariance model results in a mean estimate model identifying

the location of the channel structures, although the channels themselves have been blurred.

For the posterior realizations based on the training-image-based prior model, ρ_M^{TI}, we can compute for example the probability that a channel exists in each pixel, Figure 5c. Coincidentally this is the same as computing the point-wise average of all realizations from the posterior. This is however only the case because the prior model allows for only 0 and 1 values. Note from Figure 5c how the location of the channel is relatively sharply outlined, even though it is based on averaging all 100 posterior realizations. Thus, not only is it possible to produce realizations of the posterior with spatial features that are consistent with a relatively complex prior model, it is also clear that adding consistent prior information for this case provides significantly sharper de-blurred images.

[Fig. 4 about here.]

[Fig. 5 about here.]

Typical application of the Metropolis algorithm for image reconstruction have been based on the classical Metropolis algorithms [1,48]. This example show that extended Metropolis algorithm can be a computationally efficient used for image reconstruction using complex prior information, as quantified by the sequential Gibbs sampler. One, of many related applications, could be in PET scanning where the scanning image for a slice in the brain (the result of tomographic inversion) may be relatively smooth. A the same time one usually has a prior knowledge of how a brain might look. The method proposed has the potential to sharpen such blurred tomographic images consistent with a complex prior model.

3.2 Cross borehole tomography

We now consider a typical geophysical inverse problem in form of a non-linear cross borehole inverse problem.

Using the training image in Figure 1a, an unconditional realization, Figure 6a, is generated using the single normal equation simulation algorithm, SNESIM [45]. We consider this the reference velocity model for a synthetic cross borehole inversion problem.

An electromagnetic wave is emitted at the 20 sources located to the left in Figure 6a, and the corresponding arrival times are measured at the 40 receivers located to the right. This way 800 travel time data are recorded. 3% Gaussian noise is added to the synthetic travel-time data, Figure 6b, and used as observed data, $\mathbf{t}_{obs}$. To compute the observed travel times we use a finite difference solution to the eikonal equation [51]. There is thus a non-linear relation between data (travel time delay) and model parameters (velocities).

We now consider solving the inverse problem of inferring information about the subsurface velocity model, using the extended Metropolis algorithm to generate samples

of the posterior probability distribution, given the observed data and an assumed noise model. The likelihood of a given model, with associated estimated travel time data, $\mathbf{t}_{est}(\mathbf{m})$, is computed as

$$L(\mathbf{m}) \; = \; exp(-\frac{1}{2}(\mathbf{t}_{est}(\mathbf{m}) - \mathbf{t}_{obs})'\mathbf{C}_D^{-1}(\mathbf{t}_{est}(\mathbf{m}) - \mathbf{t}_{obs})) \quad (12)$$

where the diagonal of $\mathbf{C}_D$ is the variance of the noise added to the synthetic computed data.

[Fig. 6 about here.]

We consider 7 different prior models based on both 2-point and multiple-point random models. All prior models are assumed to have the correct mean and variance, as obtained from the training image in Figure 1a. Thus, all considered prior models are consistent with this lower order statistics of the training image.

The first 5 prior models are based on 2-point random models. A pure nugget model assumes no spatial correlation, and thus all model parameters are a-priori considered uncorrelated, ρ_M^{nugget}. Two models, $\rho_M^{Gau_1}$ and $\rho_M^{Gau_8}$, are based on a Gaussian covariance model with an isotropic range of 1m and 8m respectively. The ρ_M^{sgsim} prior is based on a covariance model inferred from the training image in Figure 1a, as an exponential covariance model with a horizontal range of 6.6m, and a vertical range of 2.2m. ρ_M^{dssim} is the same as ρ_M^{sgsim}, except that the correct binary distribution from the training image is assumed , such that pixel values can only be white or black, with a prior probability of a black pixel (indicating a channel) of 0.3, and hence a prior probability of a white pixel of 0.7. ρ_M^{TI} is based on the multiple-point statistical model inferred from the training image in Figure 1a. The last prior, $\rho_M^{TI_{90}}$, is based on the multiple-point statistical model inferred from the training image rotated 90 degrees clockwise. The 4 prior models, ρ_M^{nugget}, ρ_M^{sgsim}, ρ_M^{dssim}, and ρ_M^{TI} reflect increasing order of statistics consistent with the training image from which the reference model was generated. The other considered prior models are, to a different extent, in conflict with the actual statistical model that was used to generate the reference model.

Figures 7-8 shows 4 unconditional realizations, i.e. a sample, of the 7 considered prior models. For each of these prior models an extended Metropolis algorithm is run for 35000 iterations. The VISIM [18] and SNESIM [45] programs have been used to perform sequential simulation, needed for applying the sequential Gibbs sampler, for 2-point and multiple-point based prior models respectively. Figure 9 shows the corresponding negative log-likelihood of all considered models as a function of iteration number. Figure 10-11 show the current model at iteration 20000, 25000, 30000, and 35000.

[Fig. 7 about here.]

[Fig. 8 about here.]

The initial phase of running the extended Metropolis algorithm is called the 'burn-in' phase. This involves a random search for a location in the model parameter space where the forward responses of the models fit data within their uncertainty. One can locate the end of the burn-in process from Figure 9 as the iteration number where the log-likelihood curve flattens out. The average negative log-likelihood of realizations of the Gaussian noise model is -N/2=-400, where N=800 is the number of data. Thus, if allowed by the prior model, the log-likelihood curve should flatten out around a value of -400, indicating that sampled models fit data within their uncertainty. When the burn-in phase has been completed the algorithm has converged, and starts sampling the posterior probability distribution. Methods for determining when a Metropolis algorithm has converged can be found in e.g. [12][7][4]

[Fig. 9 about here.]

[Fig. 10 about here.]

[Fig. 11 about here.]

It is clear from Figure 9 that the burn-in phase has not been completed in the considered 35000 iterations, when using the ρ_M^{nugget} prior. The corresponding log-likelihood curve never flattens out, nor does it reach a level that indicates that the data are fitted within their uncertainty. Using $\rho_M^{Gau_1}$ the burn-in phase seems to be completed after around 20000 iterations. Note however that the associated log-likelihood level indicates that the data are relatively poorly matched. In any case, using a prior model with just a little spatial correlation, as when considering $\rho_M^{Gau_1}$, results in a sampling algorithm that is much more efficient than when considering no spatial correlation.

As the information content of the prior is increased, considering ρ_M^{nugget}, ρ_M^{sgsim}, ρ_M^{dssim}, and ρ_M^{TI}, the burn-in phase is completed increasingly faster, namely at iteration number ∞, 4000, 3800, and 1000 respectively. This indicates that the computational complexity of finishing the burn-in phase, and hence sampling the posterior probability density, is also affected by the information content of the chosen prior. An analysis of the sampling efficiency of the extended Metropolis algorithm reveals that, on average, 15000, 6000, and 2500 iterations are needed, considering the ρ_M^{sgsim}, ρ_M^{dssim}, and ρ_M^{TI} priors respectively, in order to obtain an independent realization of the posterior probability distribution. This means that using the ρ_M^{TI} prior results in a sampling algorithm that is 6 times more computational efficient than when using the ρ_M^{sgsim} prior, and about a factor of 2 more efficient than when using the ρ_M^{dssim} prior. For this example, all three prior models with spatial constraints, ρ_M^{sgsim}, ρ_M^{dssim}, and ρ_M^{TI}, perform up to infinitely more efficient that when using the ρ_M^{nugget} prior.

All the prior models inconsistent with the true model ($\rho_M^{TI_{90}}$, $\rho_M^{Gau_1}$ and $\rho_M^{Gau_8}$) perform worse that when using the ρ_M^{TI} prior. Yet, they also all perform much better than when considering the ρ_M^{nugget} prior.

Figures 7d and 10d illustrates that the use of the movie strategy, as discussed previously [46], is clearly possible using complex priors such as ones based on training images. Figure 7d shows 4 independent realizations of the . It is thus a visualization of the prior assumptions inherent in ρ_M^{TI}. Figure 10d shows a corresponding sample, in form of 4 independent realizations, of the posterior pdf, σ_{TI}, and is thus a graphical representation of the state of information of the model parameters obtained by merging the states of information as given by the prior and the likelihood, Eq. 2. The differences between Figures 7d-10d can be explained by data.

Of the seven considered prior models, it is only using the ρ_M^{TI} prior model that result in an actual sample in Figures 10-11. All other sets of models are either statistically dependent, or reflect a model at a timestep where the Metropolis algorithm has not yet converged.

This simple example suggests that an otherwise relatively easy inverse problem, such as the inversion of first arrival time data, becomes virtually unsolvable using a non-informative prior such as ρ_M^{nugget}. It however also shows that any consistent prior information will lead, not only to samples of the posterior probability distribution that are more geologically realistic, but also to a much more efficient sampling algorithm.

4 Sequential Gibbs sampling used for optimization

Simulated annealing [28] is an optimization method closely related to the Metropolis algorithm. Therefore it might seem appealing to use the sequential Gibbs sampler in combination with simulated annealing in order to locate the model with maximum posterior probability, consistent with both the prior and the data likelihood. This is, however, not readily possible.

One feature of the extended Metropolis algorithm is that an actual measure of the prior probability for a given model, $\rho_M(\mathbf{m})$, need not be explicitly computed. It is sufficient that a black box algorithm exists that perform a random walk according to the prior, which we here propose to do using sequential Gibbs sampling.

When using the simulated annealing algorithm, one must evaluate the posterior probability of a given model in each iteration. This can be done for simple Gaussian prior models [29], but there is presently no way of evaluating the prior likelihood of more complex prior models based on training images. As the prior likelihood cannot be computed, neither can the posterior probability.

If the sequential Gibbs sampler is used as part of a simulated annealing algorithm for proposing models, and the

likelihood, $L(\mathbf{m})$ used to evaluate each model, then such an algorithm would not end up in the model with maximum posterior probability. It would end up in the model with maximum data likelihood, of all possible prior models. Say the prior is given by a Gaussian model. Then all models will have non-zero prior probability. Using the simulated annealing method as described above, will then simply, in infinite time with infinite slow cooling, locate the model with maximum data likelihood, and not the maximum posterior model. The prior will in this way have no effect of the final optimization result. For a prior based on multiple point statistics, some models will have a prior probability of zero, and hence not all models will be acceptable a priori. In such a case the model located using optimization as considered above, will not necessarily be the maximum likelihood model, nor the model with maximum posterior probability, but the one model of all prior acceptable models with maximum likelihood.

Thus, using the sequential Gibbs sampler is not suited to solve optimization problems, as part of for example simulated annealing, or optimization of the posterior pdf in general, because the actual prior probability of a given model cannot be computed, which makes it imposible to locate the model with maximum posterior probability. Note that the same conclusions can be made using optimization based on the gradual deformation method [29] and the probability perturbation method [6], which are two methods that can be used to gradually change a realization of a random function based on 2-point and multiple-point geostatistics respectively. Likewise the optimization method proposed by Mariethoz et al. [30], based on a prior sampler resembling the sequential Gibbs sampler, will not locate the model with maximum posterior probability.

5 Conclusion

We have proposed an algorithm, sequential Gibbs sampling, that can be used to randomly sample a prior model described by any statistical model for which conditional realizations can be generated using sequential simulation. No analytical closed form expression of the prior pdf is needed. We have laid out the theoretical background for applying the algorithm, and shown that it is guaranteed to sample, using arbitrarily large or small step lengths, any random function that can be sampled using sequential simulation.

These properties make it ideal to use as a way to sample complex prior models as part of the extended Metropolis algorithm, allowing sampling of the posterior probability density function of non-linear inverse problems using realistic prior information.

Through two case studies we have demonstrated how samples from the posterior probability density function can honor both data and a relatively complex prior. The variability of realizations of the posterior pdf decreases as the level of consistent prior information is increased.

We have also demonstrated that the choice of a noninformative prior, such as the uniform prior, while not in conflict with the true subsurface, may cause the inverse problem to become practically unsolvable as the number of model parameters increase. In practice we have demonstrated how the inclusion of prior information increases the computational efficiency of the Metropolis sampling algorithm. As consistent prior information increases, the computational demands for generating unconditional realizations of the posterior pdf decreases. Likewise, inconsistent prior information tends to increase the computational demands for running the Metropolis sampling algorithm.

This suggest that if at all possible, an effort should be made to obtain a prior model consistent with the problem at hand. This can be achieved through for example geological expert knowledge, information from outcrops, or from other independent experiments.

State of the art geostatistical algorithms, based on sequential simulation, can already today produce realizations of quite complex random models, reflecting for example realistic geological features. There is no sign that this development will stop, and therefore the future will probably allow even more complex patterns and realistic models to be simulated. Such progress will have an immediate impact on solving inverse problems using sequential Gibbs sampling as we have discussed here.

References

1. Besag, J.: On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological) **48**(3), 259–302 (1986)
2. Bosch, M.: Lithologic tomography: From plural geophysical data to lithology estimation. Journal of Geophysical Research **104**(B1), 749–766 (1999)
3. Box, G., Tiao, G.: Bayesian Inference in Statistical Analysis. Wiley (1973)

4. Brooks, S., Gelman, A.: General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics pp. 434–455 (1998)
5. Buland, A., Omre, H.: Bayesian linearized AVO inversion. Geophysics **68**, 185 (2003)
6. Caers, J., Hoffman, T.: The probability perturbation method: A new look at bayesian inverse modeling. Mathematical Geology **38**(1), 81 – 100 (2006)
7. Cowles, M., Carlin, B.: Markov chain monte carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association pp. 883–904 (1996)
8. Deutsch, C.V., Tran, T.T., Xie, Y.: An approach to ensure histogram rerproduction in direct sequential simulation. Tech. rep., Centre for Computational Geostatistics, University of Alberta, Edmonton, Alberta (2000)
9. Fu, J., Gomez-Hernandez, J.: Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method. Inverse Problems in Science and Engineering **16**(7), 865–884 (2008)
10. Fu, J., Gómez-Hernández, J.: A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling. Mathematical Geosciences **41**(2), 105–128 (2009)
11. Gelman, A., Roberts, G., Gilks, W.: Efficient metropolis jumping rules. In: J. Bernardo, K. Berger, A. Dawid, A. Smith (eds.) Bayesian Statistics 5,, pp. 599–608. Clarendon press, Oxford. (1996)
12. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Statistical science pp. 457–472 (1992)
13. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell **6**, 721–741 (1984)
14. Gomez-Hernandez, J., Journel, A.: Joint sequential simulation of multi-Gaussian fields. Geostatistics Troia **92**, 85–94 (1993)
15. Goovaerts, P.: Geostatistics for natural resources evalutaion. Applied Geostatistics Series. Oxford University Press (1997)
16. Guardiano, F., Srivastava, R.: Multivariate geostatistics: beyond bivariate moments. Geostatistics-Troia **1**, 133–144 (1993)
17. Hansen, T.M., Journel, A.G., Tarantola, A., Mosegaard, K.: Linear inverse Gaussian theory and geostatistics. Geophysics **71**(6), R101–R111 (2006)
18. Hansen, T.M., Mosegaard, K.: VISIM: Sequential simulation for linear inverse problems. Computers and Geosciences **34**(1), 53–76 (2008)
19. Hansen, T.M., Mosegaard, K., Cordua, K.C.: Using geostatistics to describe complex a priori information for inverse problems. In: J.M. Ortiz, X. Emery (eds.) VIII International Geostatistics Congress, vol. 1, pp. 329–338. Mining Engineering Department, University of Chile (2008)
20. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97 (1970)
21. Irving, J., Singha, K.: Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resour. Res **46** (2010)
22. Jaynes, E.: Highly informative priors. In: J. Bernardo, M. DeGroot, D. Lindley, A. Smith (eds.) Bayesian Statistics 2, pp. 329–360. Elsevier (1985)
23. Jaynes, E., Bretthorst, G.: Probability theory: the logic of science. Cambridge Univ Pr (2003)
24. Journel, A., Zhang, T.: The Necessity of a Multiple-Point Prior Model. Mathematical Geology **38**(5), 591–610 (2006)
25. Journel, A.G.: Modeling uncertainty: Some conceptual thoughts. In: R. Dimitrakopoulos (ed.) Geostatistics for the Next Century, pp. 30–43. Kluwer, Dordrecht, Holland (1994)
26. Journel, A.G., Huijbregts, C.J.: Mining Geostatistics. Academic Press (1978)
27. Khan, A., Mosegaard, K.: An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data. J. Geophys. Res **107**(E6), 19–44 (2002)
28. Kirkpatrick, S., Jr., D., Vecchi, M.: Optimization by simmulated annealing. science **220**(4598), 671–680 (1983)
29. Le Ravalec-Dupin, M., Noetinger, B.: Optimization with the gradual deformation method. Mathematical Geology **34**(2), 125 – 142 (2002)
30. Mariethoz, G., Renard, P., Caers, J.: Bayesian inverse problem and optimization with iterative spatial resampling. Water Resources Research **46**(11), W11,530 (2010)
31. Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1963)
32. Mosegaard, K.: Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse problems **14**, 405 (1998)
33. Mosegaard, K.: Monte Carlo Analysis of Inverse Problems. University of Copenhagen (2006)
34. Mosegaard, K.: Quest for consistency, symmetry and simplicity - the legacy of albert tarantola. Geophysics –, – (2011)
35. Mosegaard, K., Sambridge, M.: Monte carlo analysis of inverse problems. Inverse Problems **18**(3), 29–54 (2002)
36. Mosegaard, K., Singh, S., Snyder, D., Wagner, H.: Monte carlo analysis of seismic reflector from moho and the w reflector. Journal of Geophysical Research **102**(B2), 2969–2981 (1997)
37. Mosegaard, K., Tarantola, A.: Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research **100**(B7), 12,431–12,447 (1995)
38. Mustapha, H., Dimitrakopoulos, R.: HOSIM: A high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. Computers & Geosciences (2010)
39. Oz, B., Deutsch, C.V., Tran, T.T., Xie, Y.: DSSIM-HR: a FORTRAN 90 program for direct sequential simulation with histogram reproduction. Comput. Geosci. **29**(1), 39–51 (2003). DOI http://dx.doi.org/10.1016/S0098-3004(02)00071-7
40. Remy, N., Boucher, A., Wu, J.: Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press (2008)
41. Scales, J., Sneider, R.: To Bayes or not to Bayes? Geophysics **62**(4), 1045–1046 (1997)
42. Scales, J., Tenorio, L.: Prior information and uncertainty in inverse problems. Geophysics **66**(2), 389 (2001)
43. Soares, A.: Direct sequential simulation and cosimulation. Math. Geol **33**(8), 911–926 (2001)
44. Strebelle, S.: Sequential simulation drawing structures from training images. Ph.D. thesis, Stanford University (2000)
45. Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol **34**(1), 1–20 (2002)
46. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM (2005)
47. Tarantola, A., Valette, B.: Inverse problems= quest for information. J. geophys **50**(3), 150–170 (1982)
48. Tjelmeland, H., Besag, J.: Markov Random Fields with Higher-order Interactions. Scandinavian Journal of Statistics **25**(3), 415–433 (1998)
49. Voss, P., Mosegaard, K., Gregersen, S.: The Tornquist Zone, a north east inclining lithospheric transition at the south western margin of the Baltic Shield: Revealed through a nonlinear teleseismic tomographic inversion. Tectonophysics **416**(1-4), 151–166 (2006)
50. Wu, J., Boucher, A., Zhang, T.: A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. Computers and Geosciences **34**(12), 1863–1876 (2008)
51. Zelt, C., Barton, P.: Three-dimensional seismic refraction tomography- A comparison of two methods applied to data from the Faeroe Basin. Journal of Geophysical Research **103**(B4), 7187–7210 (1998)

52. Zhang, T.: Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics. Earth Science Frontiers **15**(1), 26 – 35 (2008). DOI DOI: 10.1016/S1872-5791(08)60016-0. URL http://www.sciencedirect.com/science/article/pii/S1872579108600160
53. Zhang, T., Lu, D., Li, D.: Porous media reconstruction using a cross-section image and multiple-point geostatistics. Advanced Computer Control, International Conference on **0**, 24–29 (2009). DOI http://doi.ieeecomputersociety.org/10.1109/ICACC.2009.33

A The Gibbs Sampler with multiple parameter perturbations

Consider a Gibbs Sampler with multiple parameter perturbations, perturbing in iteration k all parameters belonging to re-simulation area U_k. Its transition probability $P_k(\mathbf{m}_i|\mathbf{m}_j)$, the probability that the algorithm in iteration k jumps to model $\mathbf{m}_i$, given that it came from $\mathbf{m}_j$, is given by:

$$P_k(\mathbf{m}_i|\mathbf{m}_j) = \begin{cases} \dfrac{p(\mathbf{m}_i)}{\sum\limits_{\mathbf{m}_k \in \mathcal{N}_j^k} p(\mathbf{m}_k)} & \mathbf{m}_i \in \mathcal{N}_j^k \\ \qquad 0 & \text{otherwise} \end{cases} \tag{13}$$

where $\mathcal{N}_j^k$ is the set of all points that are identical to $\mathbf{m}_j$ in parameters not belonging to the re-simulation area U_k, and $p(\mathbf{m})$ is the desired sampling distribution.

If $\mathbf{m}_i \in \mathcal{N}_j^k$, we can show the following symmetry property for $P_k(\mathbf{m}_i|\mathbf{m}_j)$:

$$\begin{aligned} P_k(\mathbf{m}_i|\mathbf{m}_j)p(\mathbf{m}_j)d\mathbf{m}_i d\mathbf{m}_j &= \frac{p(\mathbf{m}_i)}{\sum\limits_{\mathbf{m}_k \in \mathcal{N}_j^k} p(\mathbf{m}_k)}\, p(\mathbf{m}_j)d\mathbf{m}_i d\mathbf{m}_j \\ &= \frac{p(\mathbf{m}_j)}{\sum\limits_{\mathbf{m}_k \in \mathcal{N}_i^k} p(\mathbf{m}_k)}\, p(\mathbf{m}_i)d\mathbf{m}_i d\mathbf{m}_j \\ &= P_k(\mathbf{m}_j|\mathbf{m}_i)p(\mathbf{m}_i)d\mathbf{m}_i d\mathbf{m}_j \end{aligned}$$

where we have used that $\mathcal{N}_j^k = \mathcal{N}_i^k$. This property, called <u>detailed balance</u>, expresses that the probability of the transition $\mathbf{m}_j \to \mathbf{m}_i$ equals the probability of the reverse transition $\mathbf{m}_i \to \mathbf{m}_j$ in iteration k. Detailed balance guarantees that once the algorithm is sampling $p(\mathbf{m})$, it will continue to sample $p(\mathbf{m})$. This means that $p(\mathbf{m})$ is an equilibrium sampling distribution for the algorithm [35].

It can be shown (see, e.g., Mosegaard and Sambridge [35]) that if our transition probability distribution $P(\mathbf{x}_i|\mathbf{x}_j)$, satisfies two particular conditions (in addition to detailed balance), then $p(\mathbf{x})$ will be the only equilibrium distribution, and so the algorithm will converge towards $p(\mathbf{x})$ regardless of the starting distribution. The two conditions are:

1. <u>Aperiodicity</u>. The probability that an iteration of the algorithm results in the trivial move $\mathbf{m}_j \longrightarrow \mathbf{m}_j$ is non-zero. This is clearly satisfied by our algorithm.
2. <u>Irreducibility</u>. It is possible to go from any point $\mathbf{m}_j$ to any other point $\mathbf{m}_i$ in $\mathcal{M}$, given a sufficient number of iterations. This requirement is satisfied by our algorithm if, in K iterations, there is non-zero probability that any model parameter is perturbed.
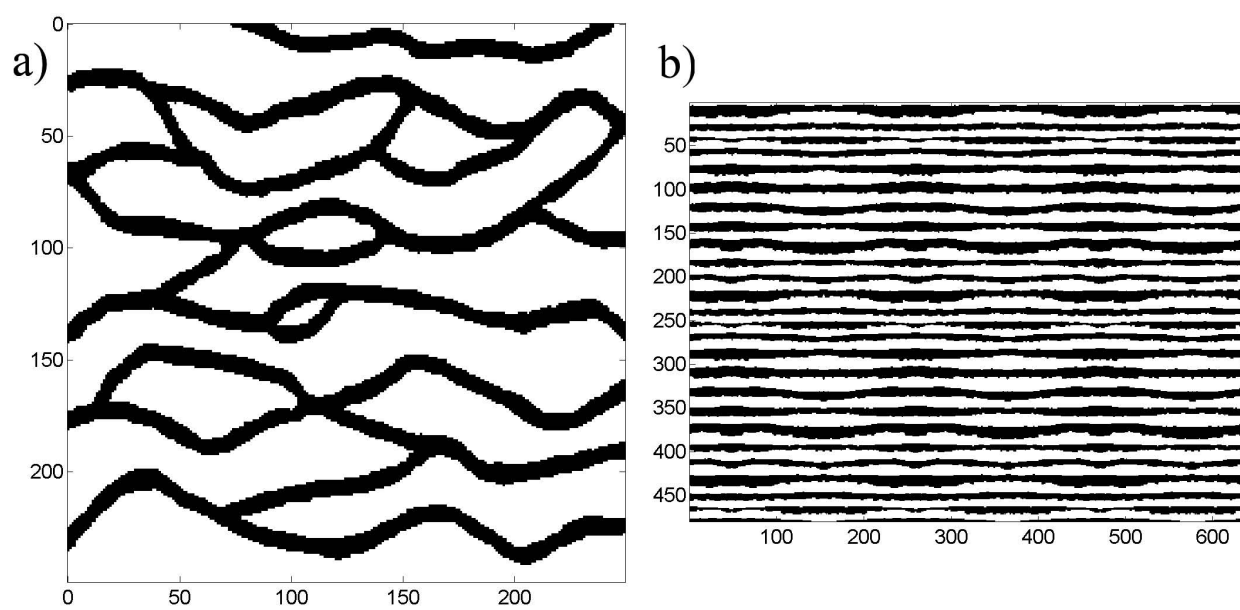
List of Figures

Fig. 1 Two different training images, reflecting a prior model based on higher order spatial moments. a) Channel based training image from Strebelle (2001). b) Pattern, used with permission from Ian Lynam.
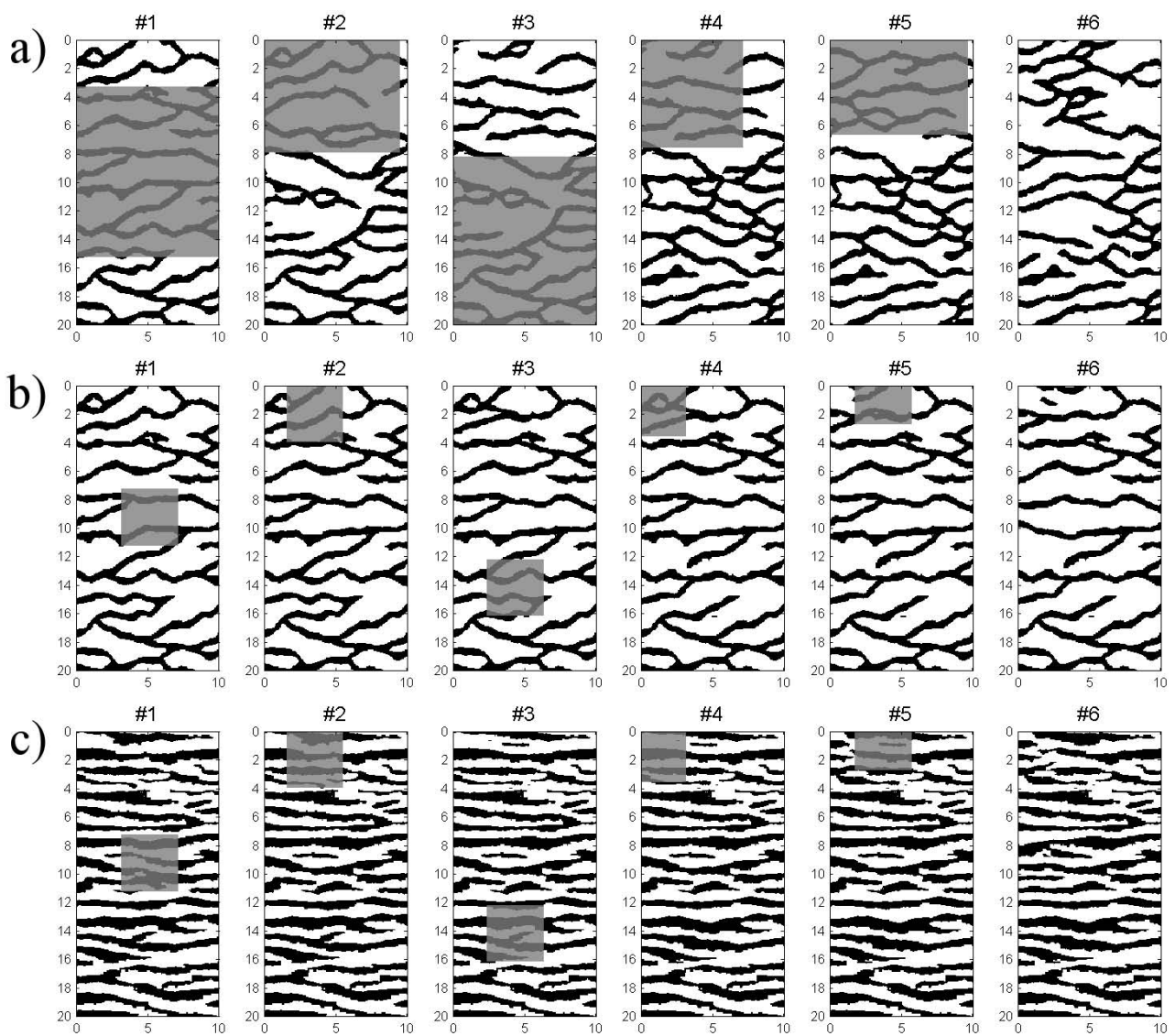
Fig. 2 Sequential Gibbs sampling using different areas of resimulation. a) 12x12 cell resimulation using the training image in Figure 1a. b) 4x4 cell resimulation using the training image in Figure 1a. c) 4x4 cell resimulation using the training image in Figure 1b. The left column is the starting model. The right column is the realization of the prior model after 6 iterations using the sequential Gibbs sampler.

a)

b)

c)

d)
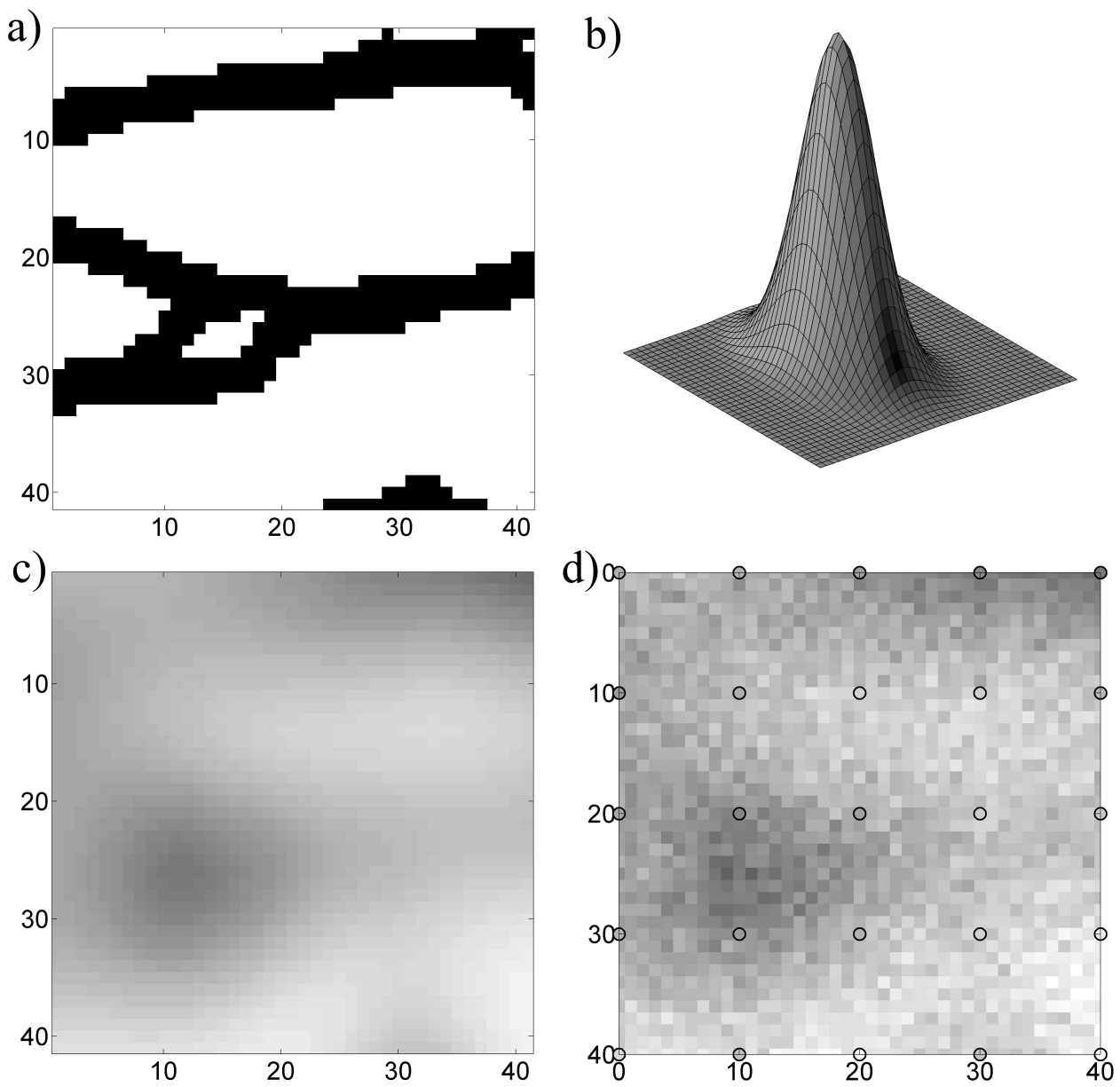
Fig. 3 a) Reference image. b) Gaussian smoothing kernel. c) Smoothed reference image. d) Smoothed reference image with noise. The 25 pixel values in the circles are used as data in the image reconstruction inversion.
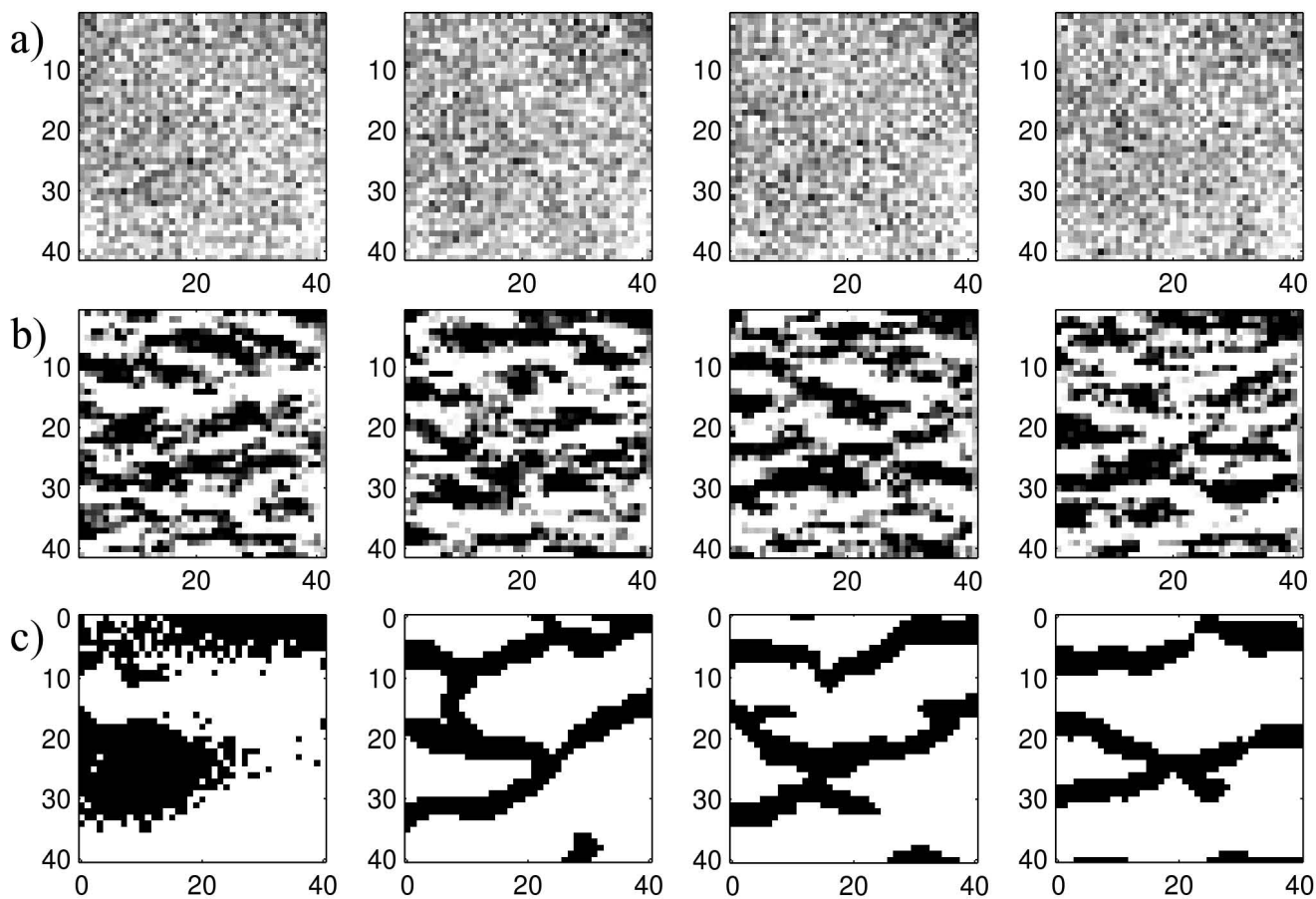
Fig. 4 posterior samples using a) ρ_M^{uncorr}, a Gaussian spatially uncorrelated prior model, b) ρ_M^{corr}, a Gaussian spatially correlated prior model (an exponential covariance model with a horizontal range of 15 pixels, and a vertical range of 6 pixels). c) initial model (leftmost) and 3 posterior realizations using the training image in Figure 1a as prior model (ρ_M^{TI}).
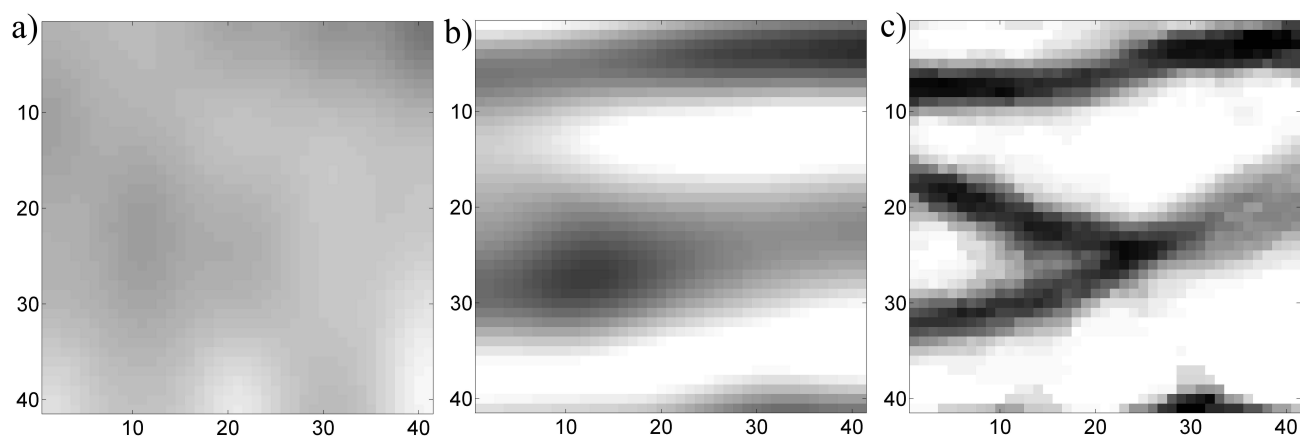
Fig. 5 posterior statistics. a) posterior mean estimate using the uncorrelated Gaussian prior model, $\rho_{\mathrm{M}}^{uncorr}$. b) posterior mean estimate using the correlated Gaussian prior model, ρ_{M}^{corr}. c) Probability of a channel structure given the use of the training image based prior model, ρ_{M}^{TI}.
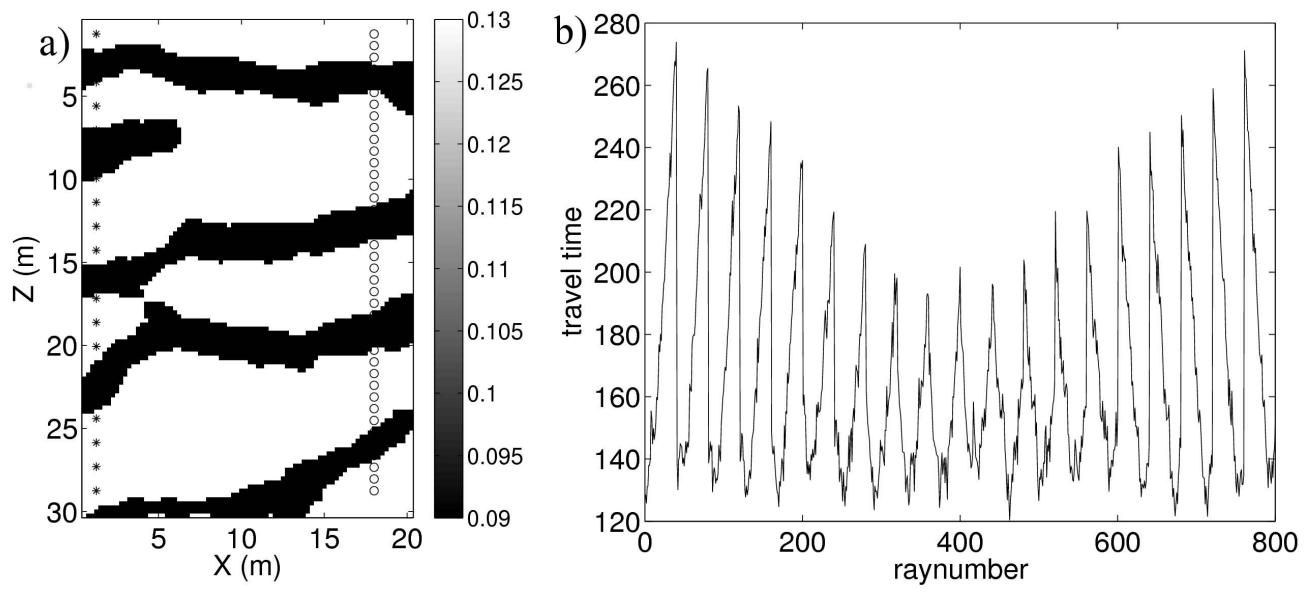
Fig. 6 a) Reference velocity model and location of sources (*) and receivers (o). Black channel structures have a velocity of 0.09 m/ns. The background velocity (white) has a velocity of 0.13 m/ns. b) Calculated first arrival travel time for waves traveling between sources and receivers. 3% normally distributed noise was added to the travel times.
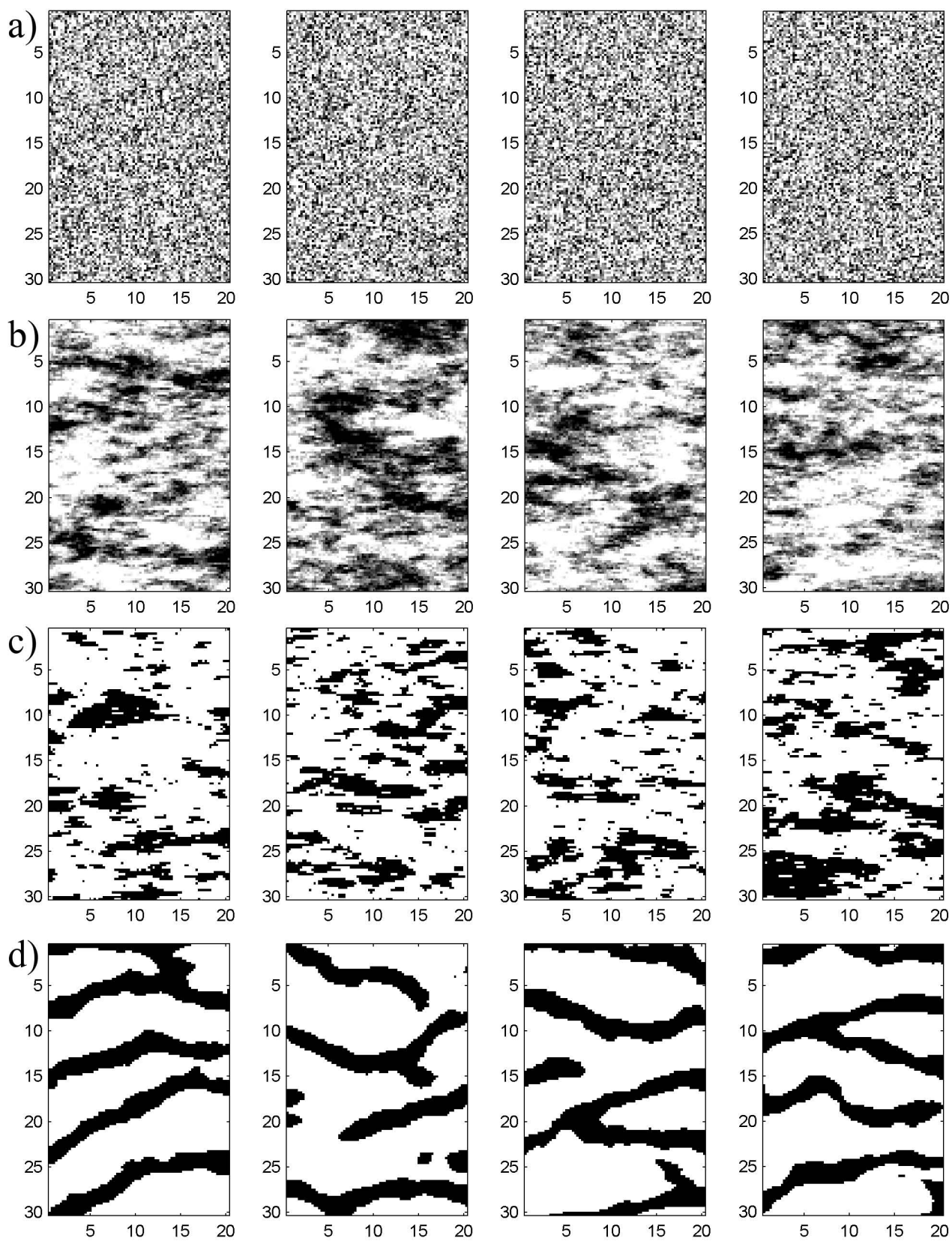
Fig. 7 4 realizations of the prior model of type a) ρ_M^{nugget}, b) ρ_M^{sgsim}, c) ρ_M^{dssim}, and d) ρ_M^{TI} prior models.
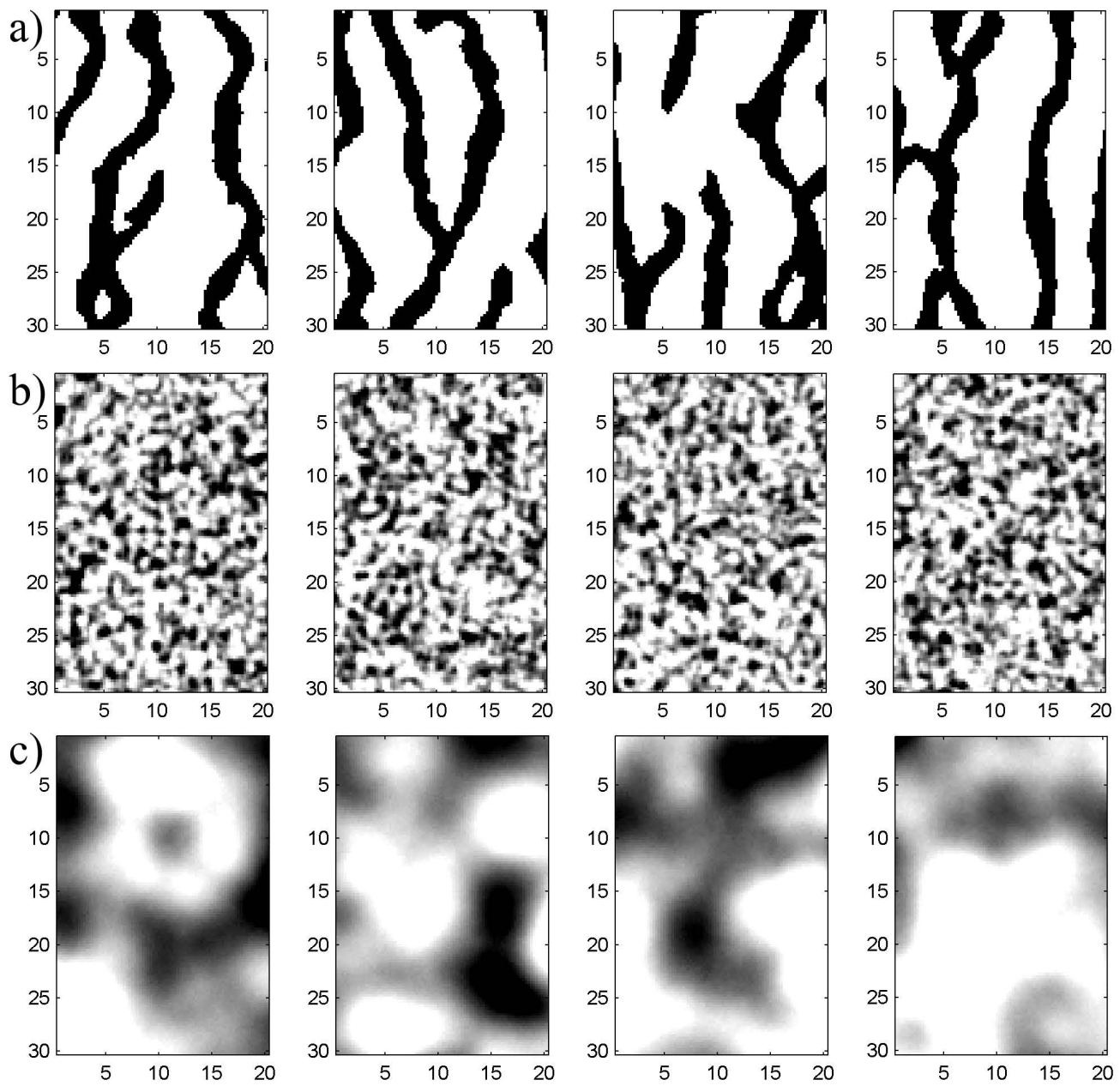
Fig. 8 4 realizations of the prior model of type a) $\rho_M^{TI_{90}}$, b) $\rho_M^{Gau_1}$, and c) $\rho_M^{Gau_8}$.
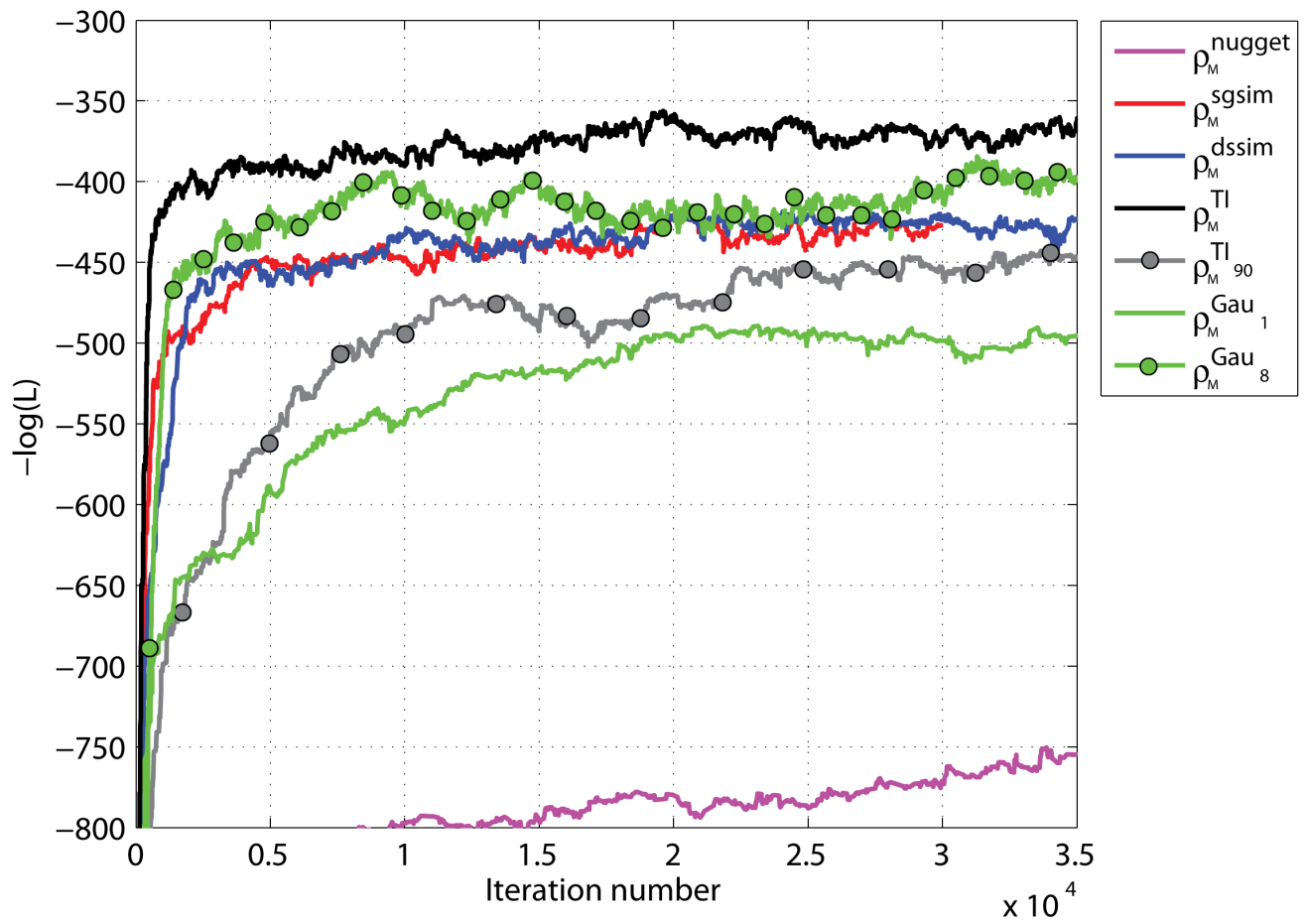
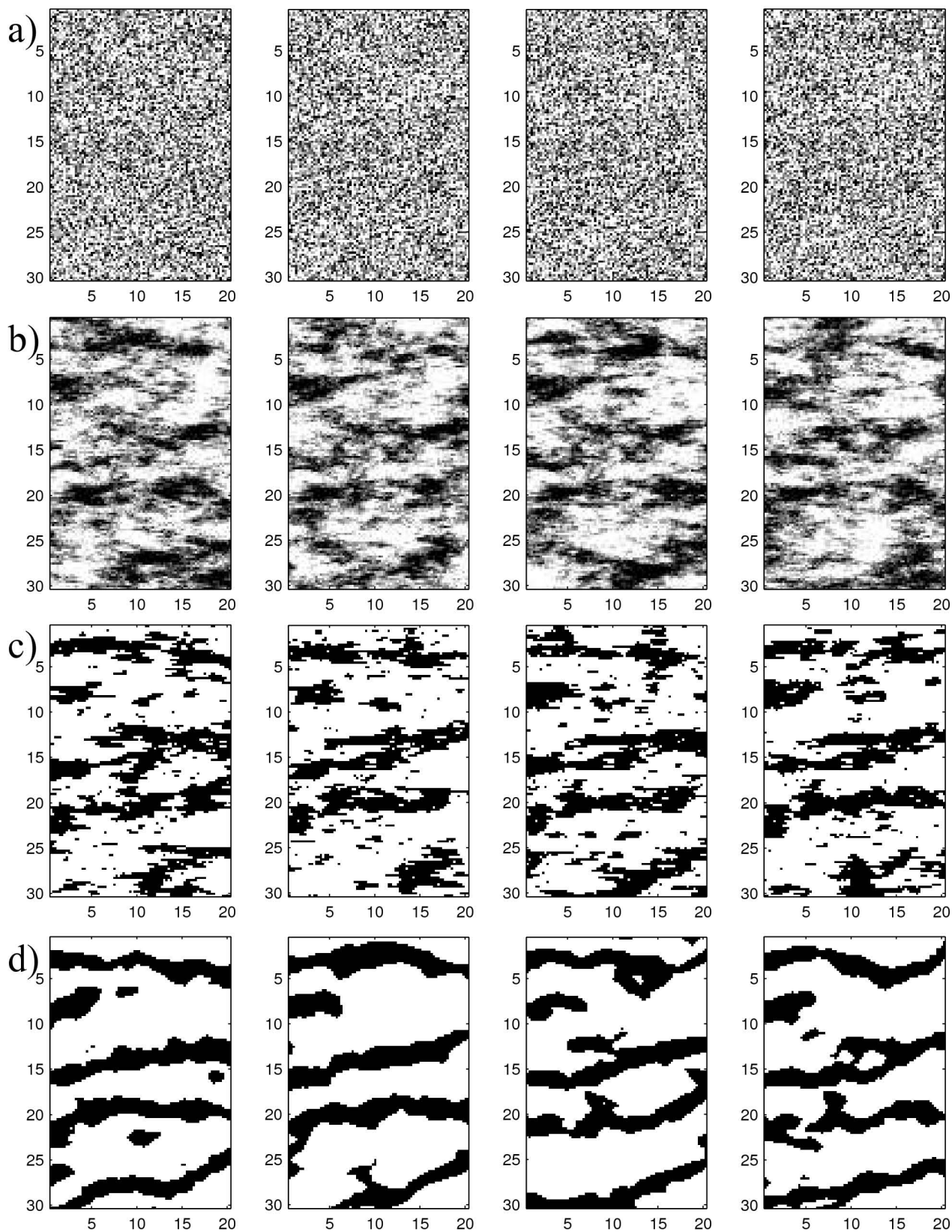**Fig. 9** Negative log-likelihood as a function of iteration number for different choices of prior model

Fig. 10 Current model at iteration number 20000, 25000, 30000 and 35000 using the a) ρ_M^{nugget}, b) ρ_M^{sgsim}, c) ρ_M^{dssim}, and d) ρ_M^{TI} prior models.
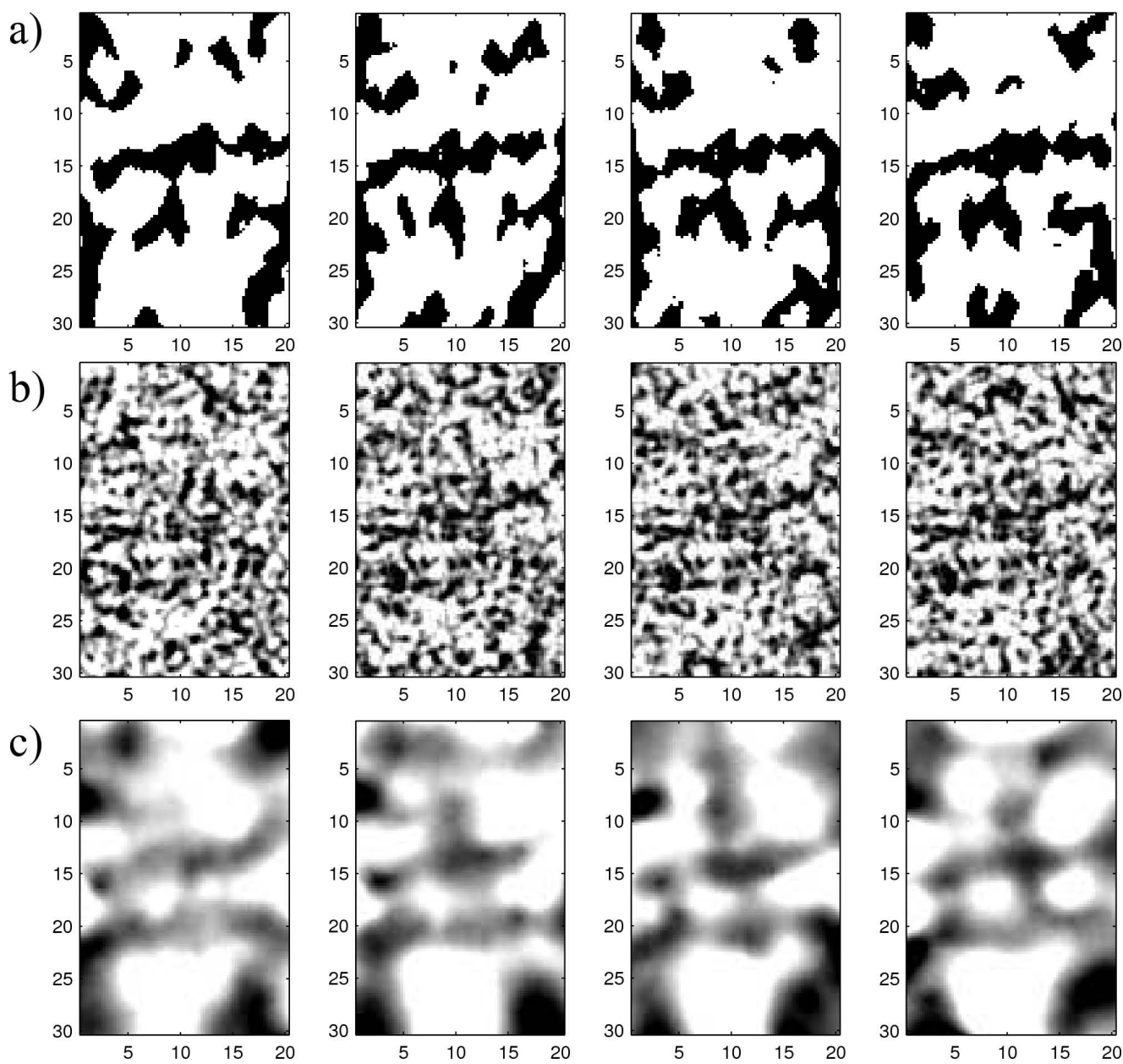
Fig. 11 Current model at iteration number 20000, 25000, 30000 and 35000 using the a) $\rho_{\mathrm{M}}^{TI_{90}}$, b) $\rho_{\mathrm{M}}^{Gau_1}$, and c) $\rho_{\mathrm{M}}^{Gau_8}$ prior models.