Technical University of Denmark

DTU

Improving multiple-point-based a priori models for inverse problems by combining Sequential Simulation with the Frequency Matching Method

Cordua, Knud Skou; Hansen, Thomas Mejer; Lange, Katrine; Frydendall, Jan; Mosegaard, Klaus

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Improving multiple-point-based a priori models for inverse problems by combining Sequential Simulation with the Frequency Matching Method

Knud S. Cordua*, Thomas M. Hansen, Katrine Lange, Jan Frydendall, Klaus Mosegaard, Technical University of Denmark, Department of Informatics and Mathematical Modelling.

Summary

In order to move beyond simplified covariance based a priori models, which are typically used for inverse problems, more complex multiple-point-based a priori models have to be considered. By means of marginal probability distributions 'learned' from a training image, sequential simulation has proven to be an efficient way of obtaining multiple realizations that honor the same multiple-point statistics as the training image. The frequency matching method provides an alternative way of formulating multiple-point-based a priori models. In this strategy the pattern frequency distributions (i.e. marginals) of the training image and a subsurface model are matched in order to obtain a solution with the same multiple-point statistics as the training image. Sequential Gibbs sampling is a simulation strategy that provides an efficient way of applying sequential simulation based algorithms as a priori information in probabilistic inverse problems. Unfortunately, when this strategy is applied with the multiple-point-based simulation algorithm SNESIM the reproducibility of training image patterns is violated. In this study we suggest to combine sequential simulation with the frequency matching method in order to improve the pattern reproducibility while maintaining the efficiency of the sequential Gibbs sampling strategy. We compare realizations of three types of a priori models. Finally, the results are exemplified through crosshole travel time tomography.

Introduction

In geostatistical and probabilistic inverse modeling, a priori models that describe the expectations of the spatial distribution of the geological structures under study are important (Journel and Zhang, 2006). Traditionally, a priori models rely on two-point statistics described through covariance models. However, such a priori models cannot capture realistically geological curvilinear structures such as tortuous channels. In order to overcome this shortcoming, multiple-point statistics has to be introduced (Guardiano and Srivastava, 1993). The Single Normal Equation SIMulation (SNESIM) algorithm is a computationally very efficient way of obtaining realizations from a joint probability density function (pdf) based on multiple-point statistics learned from a training image using sequential simulation (Strebelle, 2002).

The extended Metropolis algorithm is a general sampling algorithm that can be used to sample the solution to nonlinear inverse problems (Mosegaard and Tarantola, 1995). The extended Metropolis algorithm demands an algorithm that is able to produce perturbations between realizations from the a priori model. An efficient way of obtaining this is through sequential Gibbs sampling (Hansen et al., 2012). The extended Metropolis algorithm has previously been used in conjunction with sequential Gibbs sampling for a priori information defined through the SNESIM algorithm to sample the solution of a tomographic full waveform inverse problem (Cordua et al., 2012).

An alternative way of defining the multiple-point-based a priori pdf is the Frequency Matching Method (FMM) (Lange et al., 2011). In this approach the frequency distributions of patterns (i.e. marginal probabilities) counted in a given solution to the subsurface and in the training image are compared. By means of the Chi-square statistics, Lange et al. (2011) quantified the match between frequency distributions. In this way, they were able to jointly optimize for the a priori expectations and a tomographic dataset. Here, we define a FMM-based a priori pdf using the Dirichlet probability distribution. We show the results of sampling this distribution using the Metropolis algorithm.

When sequential Gibbs sampling is applied with the SNESIM algorithm, the reproducibility of the spatial continuity seen in the training image is reduced. This is caused by the conditional simulation technique inhered in SNESIM, which reduces the number of conditional data events when inconsistencies (i.e. singularities) occurs during the simulation. These effects are reduced for full unconditional SNESIM realizations, but are evident for the iterative perturbation strategy performed by the sequential Gibbs sampling. We suggest an a priori pdf that combines the SNESIM and FMM based a priori pdfs in order to overcome these shortcomings. We show that realizations from the combined a priori pdf ensures better reproducibility of spatial structures found in the training image than compared to the individual SNESIM and FMM-based a priori pdfs, respectively.

The importance of the reproducibility when solving inverse problems is demonstrated through a crosshole travel time tomographic inverse problem. The solution to this nonlinear inverse problem is sampled using the extended Metropolis

algorithm with both the SNESIM and the combined SNESIM-FMM-based a priori pdfs, respectively.

Methodology

Consider that the subsurface can be represented by a discrete set of model parameters $\mathbf{m}$. In geophysical inverse problems, information about the unknown model parameters is retrieved based on a set of indirect observations $\mathbf{d}$ (e.g. travel time data), a theoretical forward problem that relates model parameters and the data, and some a priori information on the model parameters. The forward relation between the model parameters and the data can be expressed as (e.g. Tarantola, 2005):

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

where g is a linear or nonlinear function that often relies on a physical law. In this study equation 1 is a nonlinear relation that provides a set of travel time data at the receiver positions given a 2D velocity field. The forward relation is based on ray-theory and is calculated using the Eikonal equation (Zelt and Barton, 1998).

In a probabilistic formulation, the solution to the inverse problem is given as an a posteriori probability density over the model parameters (e.g. Tarantola, 2005):

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m})L(\mathbf{m}), \tag{2}$$

where k is a normalization constant, $\rho_M(\mathbf{m})$ is the a priori pdf, and $L(\mathbf{m})$ is the likelihood function. $\rho_M(\mathbf{m})$ describes the probability that the model satisfies the a priori information. $L(\mathbf{m})$ describes how well the modeled data explains the observed data given a data uncertainty. Hence, the a posteriori probability density describes the combined states of information provided by the data and the a priori information.

The extended Metropolis algorithm
The extended Metropolis algorithm can be used to sample the a posteriori probability density of a general nonlinear inverse problem as formulated in equation 2. This algorithm only requires: 1) A "black box" algorithm that is able to produce perturbations between realizations from the a priori pdf. 2) An algorithm that is able to compute the likelihood for a given set of model parameters. The extended Metropolis algorithm contains the following steps:
1) The exploration step:
An a priori sampler proposes a realization, $\mathbf{m}_{propose}$, from the a priori pdf. $\mathbf{m}_{propose}$ is a perturbation of a current realization, $\mathbf{m}_{current}$.

2) The exploitation step:
The proposed realization is accepted with the probability:

$$P_{accept} = \min\left(1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{current})}\right) \tag{3}$$

If the proposed model is accepted, $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{current}$, otherwise $\mathbf{m}_{current}$ counts again.
The above procedure is continued until a desirable number of realizations have been accepted. Together, all the accepted realizations constitute a sample of the a posteriori probability density (Mosegaard and Tarantola, 1995).

Sequential Gibbs sampling
Sequential Gibbs sampling is a computationally efficient way to sample complex a priori models as quantified by most geostatistical simulation algorithms, such as for example the SNESIM algorithm (Hansen et al., 2012). With sequential Gibbs sampling the degree of perturbation between realizations can be controlled. In this way, a priori information quantified by geostatistical simulation algorithms serve as a "black box" algorithm that can be applied with the extended Metropolis algorithm to sample the solution for probabilistic inverse problems.

The flow of sequential Gibbs sampling is:
1) A current unconditional realization of the a priori pdf is provided.
2) A subset of the model parameters in the current realization is randomly chosen.
3) The model parameters within this subset are resimulated using sequential simulation conditional to the remaining model parameters (using e.g. the SNESIM algorithm).
4) Step (2) and (3) of this procedure are repeated in order to obtain multiple realizations of the a priori pdf.

The size of the subset of model parameters to be resimulated is chosen subjectively and controls the explorations nature of the Metropolis algorithm. For large subsets the exploration step becomes large and the probability of accept (in equation 3) decreases. On the other hand, smaller exploration steps leads to a higher accept probability. However, a small exploration step causes successive accepted realizations of the Metropolis algorithm to become statistically more dependent and, hence, more realizations have to be accepted to obtain statistically independent realizations. For more details on this topic see Hansen et al. (2012) and Cordua et al. (2012).

The frequency matching method
Multiple-point sample algorithms rely on sequential simulation, which is based on the fact that the complete joint probability density can be factorized by conditional

probability densities. The conditional probability densities can (according to the product rule) be expressed by means of marginal probability densities. These "marginals" are extracted (or learned) from the training image by simply counting the number of times a certain pattern occurs in image. The number of pixels within the patterns is fixed and determined by a template. The marginal pdf obtained in this way can be viewed as a frequency distribution (i.e. a normalized histogram), which is the same as the content of the search tree, as referred to by Strebelle (2002).

In the frequency matching method (Lange et al., 2011) the multiple-point-based a priori pdf is quantified by measuring the degree of fit between the frequency distribution of the training image and a current realization. In this way it becomes possible to actually quantify the multiple-point a priori pdf, which is not possible using the SNESIM algorithm.

Here, we defined the frequency matching measure using the Dirichlet pdf, which is different from the approach of Lange et al. (2011):

$$\rho_{FMM}(\mathbf{m}) = \frac{N^{cur}!}{H_1^{cur}!,...,H_K^{cur}!} \prod_{k=1}^{K} \left(\frac{H_k^{TI} + H_k^{prior}}{N^{TI} + N^{prior}} \right)^{H_k^{cur}} , \quad (4)$$

where H_k^{cur} is the number of counts in the k'th bin of the (unnormalized) histogram obtained from a current realization $\mathbf{m}$. H_k^{TI} is the number of counts in the k'th bin of the (unnormalized) histogram obtained from training image. $K = c^T$ is the number of possible pattern combinations, which is function of the template size T and the number of categories c. Further, we have that:

$$N^{cur} = \sum_{k=1}^{K} H_k^{cur} \qquad (5)$$

$$N^{TI} = \sum_{k=1}^{K} H_k^{TI} \qquad (6)$$

$$N^{prior} = \sum_{k=1}^{K} H_k^{prior} \qquad (7)$$

where H_k^{prior} is the k'th bin of the a priori (unnormalized) histogram, which represents the a priori expectation of the histogram related to underlying process before the training image histogram is observed. Hence, H_k^{prior} can be used to quantify the degree of expected match between the frequency distributions of a current subsurface image and the training image. For small values of H_k^{prior} the current model is expected to match the training image frequency distribution better than for large values. Note that the

Dirichlet distribution only needs to be evaluated for the bins $k \in \left\{ j \,|\, H_j^{cur} \neq 0 \right\}$. All other bins do not contribute to the probability. Hence, the histograms becomes sparse, which, in particular, saves memory for large template sizes and/or many categories of the model parameter values.

Combining FMM with the SNESIM algorithm
Figure 2 shows realizations from the SNESIM-based priori model using the sequential Gibbs sample strategy. Figure 3 shows realizations from the Dirichlet (i.e. FMM-based) a priori probability distribution. The multiple-point statistics of these a priori models is obtained from the training image seen in figure 1. By comparing figure 2 and 3 with the training image it is obvious that the continuous structures seen in the training image are not very well reproduced.
In order to improve this, we suggest combining the FMM with the SNESIM algorithm such that we obtained an a priori pdf defined as:

$$\rho_M(\mathbf{m}) = \rho_{SNESIM}(\mathbf{m}) \rho_{FMM}(\mathbf{m}) \qquad (8)$$

This a priori pdf can efficiently be sampled using the extended Metropolis algorithm in conjunction with sequential Gibbs sampling. By substituting $\rho_{FMM}(\mathbf{m})$ with the likelihood function $L(\mathbf{m})$ in equation (2) and (3), realizations from the combined a priori in equation 8 can be obtained. Note that, in this way, the value of $\rho_{SNESIM}(\mathbf{m})$ does not need to be evaluated.
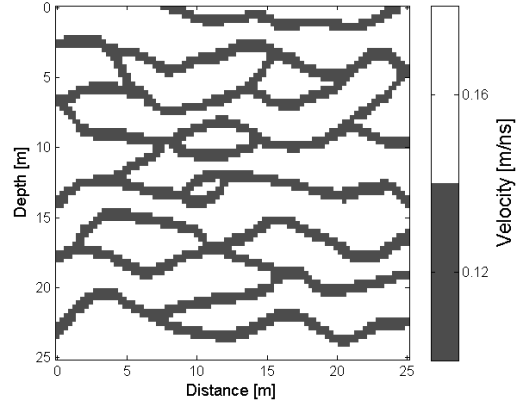

Figure 1. Training image used for obtaining the multiple-point a priori statistics.

Results

Figure 4 shows realizations obtained from the combined a priori model defined in equation 8. In this study we choose the a priori histogram to be a homogenous distribution with $H_k^{prior} = 5$, $k \in \left\{ j \,|\, H_j^{cur} \neq 0 \right\}$ and a template size of 3

pixels x 3 pixels. The results demonstrate that the combined FMM-SNESIM-based a priori probability density recovers the structures of the training image better than compared to both the SNESIM and FMM-based a priori pdfs.
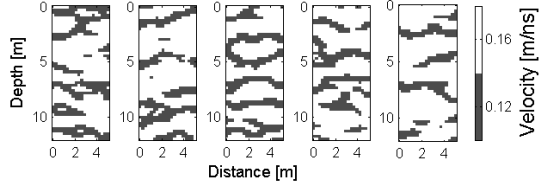

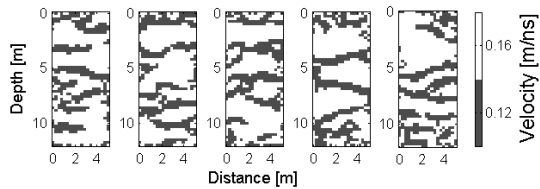Figure 2. Realizations from the SNESIM a priori model using sequential Gibbs sampling.

Figure 3. Realizations of the Dirichlet pdf (i.e. FFM-based a priori pdf) using the Metropolis algorithm with a homogenous proposal pdf.
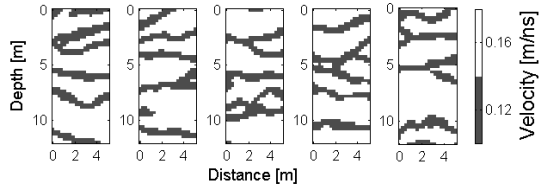

Figure 4. Realizations from the combined SNESIM-FMM-based a priori pdf using the extended Metropolis algorithm in conjunction with sequential Gibbs sampling.

Crosshole travel time tomography

In order to demonstrate how the different a priori models influence the solution to a nonlinear inverse problem, we consider a crosshole ground penetrating radar tomographic inverse problem (see e.g. Cordua et al., 2009). A synthetic reference model, from which a synthetic data set is obtained, is seen in figure 5. This model is a fully unconditional realization of the SNESIM based a priori pdf. A zero mean uncorrelated Gaussian noise component with a standard deviation of 1 ns (~2.7 % of the signal) is added to the data. The likelihood function is a Gaussian pdf that takes into account the statistics of the noise. The result of the inversion is seen in figure 6 and 7. It is clear that the improved FMM-SNESIM-based a priori probability density provides realizations that resemble the reference model better than when using the SNESIM-based a priori pdf. Moreover, the variability between the individual realizations becomes smaller when considering the

combined a priori model. This shows that the improved a priori information improves the resolution of the solution.
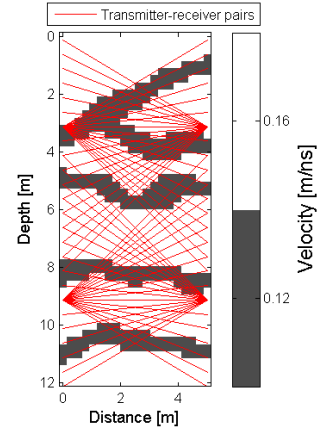

Figure 5. Reference model used for travel time tomography. The red rays give an indication of the data coverage.
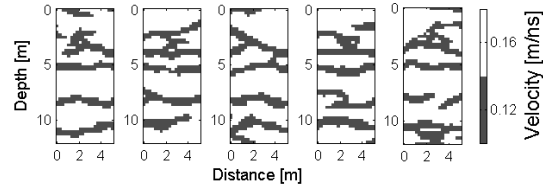

Figure 6. Realizations from the a posteriori pdf with a priori information defined by SNESIM using sequential Gibbs sampling.
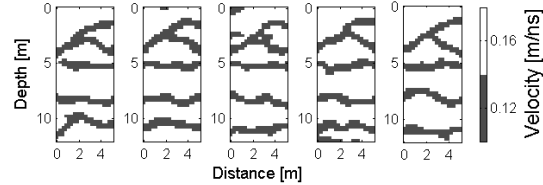

Figure 7. Realizations from the a posteriori pdf based on the combined SNESIM-FMM a priori pdf using sequential Gibbs sampling.

Discussion and Conclusion

We have demonstrated the potential of combining the FMM with the sequential simulation strategy provided by SNESIM. In this way, realizations obtained when using sequential Gibbs sampling reproduces the spatial structures of the training image much better then when only considering SNESIM. At the same time, the suggested strategy ensures that the computationally efficiency of sequential simulation is maintained.

The combined SNESIM-FMM-based a priori model demonstrates to improve the resolution when applied for a tomographic nonlinear inverse problem.

Combining Sequential Simulation with the Frequency Matching Method

References

Cordua, K. S., T. M. Hansen, and K. Mosegaard, 2012, Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information: Geophysics, **77**, H19 – H31.

Cordua, K. S., L. Nielsen, M. C. Looms, T. M. Hansen, and A. Binley, 2009, Quantifying the influence of static-like errors in least-squares-based inversion and sequential simulation of cross-borehole ground penetrating radar data: Journal of Applied Geophysics, **68**, 71 – 84.

Hansen, T. M., K. S. Cordua, and K. Mosegaard, 2012, Inverse problems with non-trivial priors: Efficient solution through Sequential Gibbs Sampling: Computational Geosciences, DOI: 10.1007/s10596-011-9271-1.

Journel, A. and T. Zhang, 2006, The Necessity of a Multiple-Point Prior Model: Mathematical Geology, **38**, 591 – 610.

Lange, K., J. Frydendall, K. S. Cordua, T. M. Hansen, Y. Melnikova, and K. Mosegaard, 2011, A Frequency Matching Method: Solving Inverse Problems by use of Geologically Realistic Prior Information: IAMG 2011, Salzburg, Austria.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of geophysical research, **100**, no. B7, 431 – 447.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, **34**, 1 – 21.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society of Industrial and Applies Mathematics, Philadelphia, PA., 353pp.

Zelt, C., and P. Barton, 1998, Three-dimensional seismic refraction seismic refraction tomography - a comparison of two methods applied to data from the Faeroe Basin: Journal of Geophysical Research, **103**, no. B4, 7187 – 7210.

Guardiano, F., and R. Srivastava, 1993, Multivariate geostatistics: Beyond bivariate moments, in A. Soares, ed., Geostatistics Tróia '92, v. 1, Kluwer, 133 – 144.