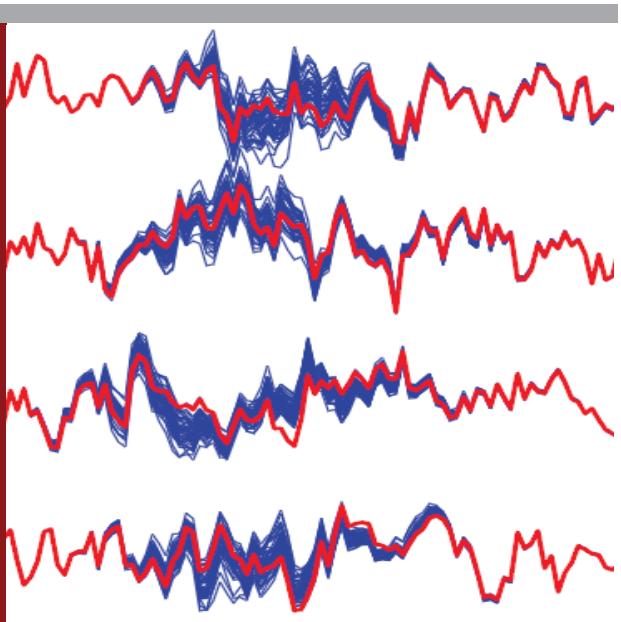
Multiple Scenario Generation of Subsurface Models

Consistent Integration of Information from Geophysical and Geological Data through Combination of Probabilistic Inverse Problem Theory and Geostatistics



Knud Skou Cordua

PhD Thesis

September 14, 2013

DTU Space
National Space Institute

Multiple Scenario Generation of Subsurface Models

Consistent Integration of Information from Geophysical and Geological Data through Combination of Probabilistic Inverse Problem Theory and Geostatistics

Knud Skou Cordua

PhD thesis

September 2013

DTU Space, National Space Institute

Division of Mathematical and Computational Geoscience

Technical University of Denmark

Kongens Lyngby, Denmark

Dedicated to those I love and care about

Table of Contents

Preface

This PhD thesis entitled "Multiple Scenario Generation of Subsurface Models - Consistent Integration of Information from Geophysical and Geological Data through Combination of Probabilistic Inverse Problem Theory and Geostatistics" has been submitted as a partial fulfillment of the requirements for a Ph.D. degree at the Technical University of Denmark (DTU). The work is based on three years research carried out at DTU Informatics at the department of Scientific Computing, DTU Space at the department of Mathematical and Computational Geoscience, and DTU's Center for Energy Resources Engineering (CERE). The work has been carried out under supervision of my main supervisor, Klaus Mosegaard and co-supervisor, Ida Lykke Fabricius. The project has been financed by DONG E&P. During the PhD project, I was four months at Stanford Center for Reservoir Forecasting (SCRF) at the Department of Energy Resources Engineering, Stanford University, USA. Here, I worked with associate professor, Jef Caers and other people from the department.

The PhD thesis consists of a summary of the scientific developments within the field of probabilistic inverse problem theory and geostatistics until now and a collection of research papers. Especially, the focus is directed at the resent years of research related to the merging of these two, hitherto separated, fields of research. My point of departure for this summary is 15 publications that have been published in or submitted to internationally peer-reviewed scientific journals or conference proceedings in the period 2010 – 2013. The collection of research papers consists of five published peer-reviewed journal articles (one first author and four co-authored), three papers submitted to peer-reviewed journals (two first author and one co-authored), and seven published conference proceedings/extended abstracts (four first author and three co-authored). These papers are found in the appendix of this thesis. Some of these papers will be referred to in the summary as paper A1, paper A2, etc. (e.g., Cordua et al., 2012 (paper A4)). These papers represent my scientific contributions to this Ph.D. work. See the introduction of the appendix section for more details.

Kongens Lyngby, September 14, 2013

Knud Skou Cordua

Acknowledgements

I address special thanks to my supervisor Klaus Mosegaard and my unofficial supervisor Thomas Mejer Hansen for their great enthusiasm, for always being in a good mood, and for interesting and very educational discussions. Also, I would like to thank Katrine Lange, Yulia Melnikova, Ida Lykke Fabricius, Angeliki Xenaki, Bo Holm Jacobsen, Jakob Heide Jørgensen, Jan Frydendall, and Jef Caers for fruitful discussions and collaborations. I would like to thank my colleagues at DTU for a great time. In particular, I am grateful for the cozy lunches with Curry Herrings and other necessary accessories. I wish to thank my wife, friends, and family for their great support during these years.

Finally, thanks to our sponsors DONG E/P for their financial support to this project.

Summary

In geosciences, as well as in astrophysics, direct observations of a studied physical system and objects are not always accessible. Instead, indirect observations have to be used in order to obtain information about the unknown system, which leads to an inverse problem. Such geoscientific inverse problems face the challenge of determining a set of unknown model parameters based on a set of indirect observations of the subsurface.

In a traditional least-squares formulation of the solution to an inverse problem, a subjectively chosen regularization parameter is used to obtain a unique solution to this problem, which leads to a smooth solution with no geological realism. Moreover, such a optimization-based framework does not allow introducing realistic geological prior information (due to a vectorial normed space structure).

This thesis focuses on a more sophisticated approach based on a probabilistic formulation of the solution to the inverse problem. In this formulation, different sources of information about the subsurface can be weighted with regard to their relative quality and reliability (i.e., uncertainties) using probability distributions and subsequently integrated into a posterior probability distribution over the model parameters. The different sources of information are provided in form of a set of observed data, uncertainties related to the data, and geological prior information, which is established from, e.g., expert knowledge and old data sets.

The prior information, when being informative and realistic, has a regulating effect on the solution to the inverse problem as geological and geophysical information are orthogonal in some ways, which allows reducing the underdetermination of the inverse problem. At the same time, such prior information also reduces the effective dimension of the inverse problem, which may considerably reduce the computationally cost related to such problems. Moreover, the probabilistic formulation of the inverse problem allows the use of geologically more realistic prior information that leads to solutions to the inverse problem with a higher degree of geological realism. Finally, the probabilistic formulation provides a means of analyzing uncertainties and potential multiple-scenario solutions to be used for risk assessments in relation to, e.g., reservoir characterization and forecasting.

Prior models rely on information from old data sets or expert knowledge in form of, e.g., training images that expresses structural, lithological, or textural features. Statistics obtained from these types of observations will be referred to as sample models. Geostatistical sampling algorithms use a sample model as input and produce multiple realizations of the model

parameters that, to some degree, honor this information. Such algorithms can be used to define the prior information for probabilistic inverse problems. In this way, very informative and geologically more realistic prior information can be provided.

This thesis provides an overview of the scientific developments within the fields of probabilistic inverse problem theory and geostatistics, with emphasis on the combination of these scientific disciplines. In particular, the focus will be on consistent probabilistic formulations of this problem, which means that a correct weighting of the different sources of information is obeyed such that no unknown assumptions and biases influence the solution to the inverse problem. This involves a definition of the probabilistically formulated inverse problem, a discussion about how prior models can be established based on statistical information from sample models, and an analysis of geostatistical algorithms in order to understand the implicit assumptions made by such “black box” algorithms.

A description of the posterior distribution can be obtained by drawing a representative sample from this distribution. Methodologies to be used for this purpose are presented. An example of sampling the posterior probability distribution of a computationally hard full waveform inverse problem using prior information based on multiple-point statistics, obtained from a training image, is demonstrated.

For some computationally challenging inverse problems, a sample from the posterior distribution might still be too laborious to be obtained. Instead, a set of model parameters with (near) maximum posterior probability can be obtained. In order to do this, a closed form mathematical formulation of the prior probability distribution has to be established, such that the posterior probability distribution can be evaluated. Different solutions to this problem are presented and discussed.

The prior probability distribution that is sampled by geostatistical sampling algorithms is typically unknown or sometime only a part of or an approximation to the distribution is known. This thesis provides an analysis and a discussion of how these prior probability distributions can be established, such that it is consistent with information provided by a known sample model. It is described how assumptions about the distribution, in addition to the information provided by the sample model, have to be made in order to end up with a unique solution to this problem. It is shown that these sampling algorithms typically provide samples from a prior probability distribution that is not consistent with the sample model. However, examples of consistent algorithms are also provided.

A likelihood function is part of the probabilistic formulation of the inverse problem. This function is based on an uncertainty model that describes the uncertainties related to the observed data. In a similar way, a formulation of the prior probability distribution that takes into account uncertainties related to the sample model statistics is formulated.

Prior models that are consistent with the statistics from a training image do not necessarily produce realizations with the same spatial patterns as seen in the training image because the local Markov properties that is satisfied in this way does not lead to a global reproduction of the pattern distribution from the training image. A prior probability distribution, with realizations that resemble the patterns as seen in the training image, is described and an efficient sampling algorithm that samples this distribution is provided. Moreover, an example of using this prior model for an inverse problem is demonstrated.

The theoretical forward problem that describes the relation between data and model parameters is often associated with some degree of approximation. This approximation may have a great impact on the solution to the inverse problem because such approximate calculations of the data have an impact similar to observation uncertainties. We refer to the effect of these approximations as modeling errors. Examples that show how the modeling error is estimated are provided. Moreover, it is shown how these effects can be taken into account in the formulation of the posterior probability distribution.

Common to the methods and strategies presented in this thesis is that they strive for a solution to the inverse problem that is consistent with the available information and to a less degree based on unconscious or subjective choices and implicit assumptions. Future studies related to theoretical developments of these strategies have to be provided. Moreover, applications of these strategies will reveal the practical implications of these consistent formulations. This will in particular be of great importance when it comes to assessments related cases of high risk such as human health or resources of high economical potentials.

1 Introduction to inverse problems and geostatistics

Characterization of the properties and structures of the subsurface is of great importance in relation to 1) protection of drinking water resources, 2) exploration and exploitation of oil, gas, and water reservoirs, 3) reservoir production and forecasting, and 4) scientific research of geological processes and dynamics. Geoscientific inverse problems deal with the problem of inferring information about the subsurface based on some indirect observations of subsurface properties.

Inverse problems occur in many aspects within geosciences. Examples of such problems are inversion of tomographic data such as ground penetrating radar (GPR) travel time data (Cordua et al., 2008), GPR amplitude data (Buursink et al., 2008), GPR full waveform data (Ernst et al., 2007; Cordua et al., 2012 (paper A4)), electricity data (Binley et al., 2002), seismic travel time data (Lange et al., 2012 (paper A9)), seismic full waveform data (Belina et al., 2009). Other examples of typical problems are inversion of reflection seismic data (Gauthier et al., 1986; Buland and Omre, 2003; Barros et al., 2010), production data (Caers, 2003), and tracer concentration data (Irving and Singha, 2010).

The common goal of inverse problems is to infer information about a set of model parameters $\mathbf{m} = \left(m_1, m_2, ..., m_N\right)^T$ that describe some property of the subsurface model being investigated based on a set of indirect observations $\mathbf{d}^{obs} = \left(d_1^{obs}, d_2^{obs}, ..., d_M^{obs}\right)^T$ of the system. These model parameters are typically related to spatial positions in a regular grid of cells (or pixels) such that the values of these model parameters is an (1D, 2D, or 3D) image of the subsurface. This will also be the case throughout this thesis.

1.1 Regularization based inversion strategies

In linear or linearizable inverse problems, the relation between calculated data $\mathbf{d}$ and a set of model parameters is given as

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \tag{1}$$

where $\mathbf{G}$ is a linear mapping operator. This relation often relies on a physical rule. If the problem of determining a set of unknown model parameters based on a set of observed data is underdetermined, the inverse of $\mathbf{G}$ does not exist. A unique solution to this problem can

be obtained using regularization. A simple solution to this problem is given by Tikhonov (1963) as

$$\tilde{\mathbf{m}} = \left(\mathbf{G}^T\mathbf{G} + \theta\mathbf{I}\right)^{-1}\mathbf{G}^T\mathbf{d}^{obs}, \quad (2)$$

where $\mathbf{I}$ is an identity matrix and θ is a positive regularization parameter.

In this case, the solution to the inverse problem depends on the regularization parameter. Roughly speaking, the value of this parameter controls the smoothness of the solution to the inverse problem and the value is chosen subjectively such that the solution becomes visually acceptable.

The regularization parameter imposes prior information into the inverse problem in a subjective way, which makes the information content and the relative weight between the observed data and the prior information opaque. Moreover, this formulation does not provide any uncertainty of the estimated solution.

1.2 Pioneering work on probabilistically formulated inverse problems

Tarantola (1982a) suggested a probabilistic formulation of the solution to the inverse problem. In this formulation, prior states of information about the model parameters $\rho_M(\mathbf{m})$ and the data $\rho_D(\mathbf{d})$ are given as probability distributions and are assumed independent.

Given a linear or non-linear forward relation between the calculated data and the model parameters

$$\mathbf{d} = g(\mathbf{m}), \quad (3)$$

the solution to the probabilistically formulated inverse problem is given as the posterior probability distribution (Tarantola, 2005)

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m})\rho_D(g(\mathbf{m})) = k\rho_M(\mathbf{m})L(\mathbf{m}) \quad (4)$$

Here k is a normalization constant, $L(\mathbf{m})$ is the likelihood function that describes how well the observed data fits the calculated data. This function is based on an uncertainty model $\rho_D(\mathbf{d})$ that describes the uncertainties related to the observed data (i.e. the prior state of information on the data) and a forward relation between the data and the mode parameters. $\rho_M(\mathbf{m})$ contains the prior information of the model parameters and is often referred to as

the prior model. Using the probabilistic formulation, the relative weighting of the different sources of information stemming from the observed data and the prior information is defined in an explicit way through the individual probability distributions.

This explicit formulation paves a way for defining a posterior probability distribution (i.e., a solution to the inverse problem) that is consistent with the observed data and prior states of information and not based on some subjectively chosen regularization parameter.

Using the probabilistic formulation of the inverse problem, a better understanding of the Tikhonov regularization can be obtained. The Tikhonov regularization in equation 2 turns out to be a maximum posterior solution to the probabilistic formulation in equation 4 in the special case where the forward relation is linear (i.e., given by equation 1) and the prior states of information for both the model parameters and the data are described by a zero-mean isotropic Gaussian distribution. In this case the regularization parameter is given as

$$\theta = \frac{\sigma_d^2}{\sigma_m^2}, \tag{5}$$

where σ_d is the standard deviation of the data uncertainties and σ_m is the a priori expected standard deviation of the model parameters (around an expectation of zero).

Tarantola and Valette (1982b) used the probabilistic formulation in equation 4 for linear and weakly non-linear inverse problems with Gaussian distributed noise and prior information, which leads to a Gaussian or close to Gaussian posterior distribution. For these cases, an expression for an iterative optimization procedure that searches for the set of model parameters with maximum posterior probability was provided.

A very challenging geophysical non-linear inverse problem is the inversion of full waveform data. Tarantola (1984) suggested combining the optimization procedure for the probabilistically formulated full waveform inverse problem with the adjoint-based method in order to obtain a computationally feasible inversion procedure for these kinds of problems. Tarantola (1986, 1988) further developed this approach for the elastic and viscoelastic wave equations. Developments and experiments based on these waveform inversion strategies are still going on (Barnes et al., 2008; Ernst et al., 2007; Yang et al., 2013).

For the highly non-linear inverse problem of estimating the static errors in reflection seismic problems, Rothman (1985, 1986) found a set of model parameters with near maximum posterior probability using the stochastic optimization simulated annealing.

However, for these non-linear inverse problems only a set of model parameters with near maximum posterior probability without any uncertainty estimate was obtained.

1.3 Sampling the solution to inverse problems

For non-linear inverse problems, the posterior distribution typically becomes an (unknown) non-Gaussian distribution (Mosegaard, 2006), which may have several areas of models (i.e. sets of model parameters) with significant probability separated with areas of models with low posterior probability. In such a case, it may be insufficient to come up with just a single set of model parameters with high posterior probability. Therefore, Mosegaard and Tarantola (1995) suggested sampling the posterior distribution related to the probabilistic formulation of the inverse problem. In this way, not only a set of model parameters with (near) maximum posterior probability was obtained, but instead a characterization of the shape of the posterior distribution, in form of a representative sample, was the target.

From such a sample, the uncertainty around the maximum posterior solution, or possibly around several considerably different solutions (in case of multimodal shaped posterior distributions), can be obtained (Mosegaard, 1998). Such ambiguities typically arise as a consequence of the inverse problem either being underdetermined (i.e. different sets of model parameters explain the same data set) and due to the uncertainties related to the observed data (described by the data error model). The degree of variability of a sample from the posterior distribution is influenced by the amount of information provided from the data, the uncertainty model, and the prior model.

Mosegaard and Tarantola (1995) proposed to sample the posterior distribution of the probabilistically formulated inverse problem using an extended edition of the Metropolis algorithm (Metropolis et al., 1953). In this extended Metropolis algorithm, no closed form mathematical expression for the prior information is needed. An algorithm that is able to sample the prior distribution is sufficient. In this way, a sample from the posterior distribution of a general non-linear inverse problem with an arbitrary prior distribution can be obtained.

Roughly speaking, the extended Metropolis algorithm has two phases (e.g. Mosegaard and Sambridge, 2002). In the first phase, known as the burn-in period, the algorithm searches for an “area of models” with significant posterior probability. In the second stage, after burn-in, the algorithm starts to sample the area of significant posterior probability. That is, the first part can be considered almost as a pure optimization procedure and the second part is the actual sampling. Hence, sampling of the posterior distribution will always be computationally harder than only searching for a set of model parameters with (near) maximum posterior

probability. For some problems, it might even be computationally prohibited to go for a sample of the posterior distribution.

Examples of using the extended Metropolis algorithm to sample the posterior distribution of a probabilistically formulated inverse problem are seen in for example Dahl-Jensen et al. (1998), Khan et al. (2000), and Voss et al. (2006). However, in these examples, as well as in earlier examples of using probabilistic inversion, only very simple prior models were applied.

During the last two decades, more complex and geologically more realistic prior models have been developed within the geostatistical society. Therefore, as noticed by Mosegaard (2006), recent advances in the geostatistical society may provide methods that could be integrated into the inversion.

1.4 Geostatistically formulated prior information

Geostatistical modeling basically rely on two different approaches: (1) the object-based approach, in which the geometrical shapes of geobodies are defined by stochastic parameters (e.g. Deutsch and Wang, 1996), and (2) the pixel-based approach, in which the earth model is characterized by an image where the pixel values are associated with stochastic parameters (e.g. Journel and Isaaks, 1984). In this thesis, only the pixel based geostatistical models will be considered.

Geostatistics was traditionally developed for probabilistic predictions of ore grades (Krige, 1951). Later on, the methodology was used for understanding porous media (Matheron, 1967) and was further developed for uncertainty estimation in a reservoir (Delfiner and Chiles, 1977). Since then, various geostatistical algorithms, such as indicator simulation (Journel and Alabert, 1989), truncated Gaussian simulation (Matheron et al., 1987), sequential Gaussian simulation (SGSIM), direct sequential simulation (DSSIM) (Deutsch and Journel, 1998), and single normal equation simulation (SNESIM) have been developed and used for various purposes related to uncertainty estimation in the earth sciences related to, e.g., production data (Caers and Hoffman, 2006) and satellite data (Boucher et al., 2008).

Like for the first applications of the probabilistically formulated inverse problems, the first geostatistical algorithms also relied on a covariance based formulation. This approach has later on been referred to as two-point statistics because it only considers pairs-wise correlations between the model parameters. Two-point statistical approaches are inadequate

when the geological prior information describes complex structures, such as bended and curved formations, which can only be captured if correlation between more than two model parameters is considered (Journel and Zhang, 2006).

Therefore, more complex and geologically more realistic geostatistical prior models, based on multiple-point statistics, have been developed (Guardiano and Srivastava, 1993; Strebelle, 2002; Zhang et al., 2006). In this approach, the correlation between more than two model parameters is considered. Unlike the covariance based prior models, the prior models described by multiple-point based algorithms are non-parametric and based on multiple-point statistics obtained from a so-called training image (Wang, 1996). The training image is a conceptual image of the subsurface, potentially provided by a geologist who has some prior knowledge about the subsurface.

Several pixel-based geostatistical sample algorithms that can be used to generate multiple realizations[1] that, to some degree, honor the multiple-point statistics from a training image, have been developed. SNESIM (Stebelle, 2002) was the first example of a computationally fast implementation of a multiple-point based sequential simulation algorithm that was based on the original idea by Guardiano and Srivastava (1993). Other examples of such geostatistical algorithms are HOSIM (Mustapha and Dimitrakopoulos, 2010), The Direct Sampling Method (Mariethoz, 2010b), and DISTPAT (Honarkhah, 2011). Some multiple-point based sampling algorithms use graphic-based approaches where the patterns are patched together. Examples of such algorithms are SIMPAT (Arpat, 2005; Arpat and Caers, 2007), CCSIM (Tahmasebi, 2012), and Filtersim (Zhang, 2006; Wu et al., 2007).

[1] Having defined a probability distribution over a set of model parameters, a realization refers to a realization of these model parameters from this distribution. A set of multiple realizations of the model parameters constitute a representative sample from this probability distribution.

2 Bridging the gap between geostatistics and probabilistic inverse problem theory

Geostatistical models are useful tools for quantifying geological information in a probabilistic manner. This provides the opportunity to use prior information described by such models in the probabilistically formulated inverse problem. Informative prior information typically implies a high degree of spatial autocorrelation among the model parameters, which is related to the expectations about the geological structures in the subsurface (Remy et al., 2009).

Prior information may serve the inverse problem in two ways (Barnes et al., 1996):

1) It provides geological prior information about the subsurface properties to be inverted for, which leads to more realistic solutions.

2) It reduces the effective dimension of the inverse problem, which makes it computationally easier to solve, in particularly when using a Monte Carlo strategy for sampling the posterior distribution (Hansen et al., 2009).

Moreover, incorporation of more information into the problem, either in form of more observations, less uncertain observations, or informative prior information, leads to less uncertainty in the solutions (i.e., less variation in a sample from the posterior distribution).

2.1 Gradual deformation and probability perturbation method

The gradual deformation method (GDM) is used to gradually move from one realization of a multi-Gaussian (prior) distribution to another realization (Hu, 2000; Hu, 2002). Le Ravalec-Dupin and Noetinger (2002) used this method to perturb the Gaussian prior model until a set of calculated production data matched a set of observed data. In this way, they found a set of model parameters that jointly honored (i.e., was a realization from) the Gaussian prior model and the observed data.

Another example of a prior perturbation strategy that is based on multiple-point statistics is the probability perturbation method (PPM) (Caers and Hoffman, 2006). This method uses the $\tau-$model (Journel, 2002) to obtain perturbations between two realizations from a multiple-point based sampling algorithm. This method has been used to optimize for a set of

model parameters that jointly honor the multiple-point statistics from a training image and a production data set (Caers, 2003; Hoffman and Caers, 2004).

The essential part of both GDM and PPM is the possibility of performing a controllable degree of perturbation (i.e., step-length) between two successive realizations from the prior model. This perturbation strategy is used to design a stochastic optimization procedure where the forward (e.g., flow) simulator is evaluated for each prior realization. This procedure continues in an iterative way until a satisfying data fit is obtained.

In order to make this optimization procedure efficient, an appropriate perturbation (i.e., step-length) between the individual prior realizations has to be chosen. If the step-length is too small, too many forward simulations have to be evaluated and the algorithm will perform very inefficient. A too large step-length may result in a lot of iterations that do not lead to an improved data fit.

The applications of the GDM and PPM are some of the first attempts of combining geostatistically based prior information with inverse problems.

2.2 Lack of consistency

A set of model parameters that are optimized using for example GDM or PPM will somehow honor both the prior model and the observed data. However, such a set of model parameters will not have maximum posterior probability (as it is formulated in equation 4), but instead be a realization of the prior model with highest possible likelihood.

Caers and Hoffmann (2006) and Suzuki and Caers (2008) suggest running multiple optimizations with different initial guesses using these methods in order to obtain a sample from the posterior distribution. However, even though multiple realizations obtained in this way will honor both the prior statistics and the observed data, they will not be consistent with the posterior distribution (i.e., not be distributed according to the posterior distribution).

Mariethoz et al. (2010a) formulate, in a probabilistic way, an inverse problem with prior information based on multiple-point statistics described by a geostatistical algorithm. They suggest using the extended Metropolis algorithm for optimization for high posterior probability with respect to the model parameters by only accepting improvements in the likelihood values. However, like in the case of GDM and PPM, this strategy will not search toward maximum posterior solutions, but instead prior realizations with maximum likelihood.

Moreover, Mariethoz et al. (2010a) suggested approximating a sample from the posterior distribution by running multiple optimizations using their extended Metropolis approach. A somehow optimized set of model parameters that are also a realization of the prior distribution is used as an input prior realization for the rejection sampler. The realizations accepted in this process are considered as realizations from the posterior distribution. However, as it is also noticed by the authors, this strategy only leads to an approximation to the posterior sample because the input "prior realizations" that are evaluated by the rejection sampler are not realizations from the data independent prior distribution, but influenced by the observed data. A pure application of the rejection sampler, in which the prior realizations were not dependent on the data, would have been computationally impossible (Hansen et al., 2013 (paper A14)).

2.3 Bridging the gap in a consistent way

A consistent combination of probabilistic inverse problem theory and geostatistics for covariance based linear inverse problems was first described by Hansen et al. (2006). This formulation relies on the probabilistic formulation as defined in equation 4. In this work, both a sample from and an optimum of the posterior distribution can be obtained (Hansen et al., 2008b). See Cordua et al. (2009) for an example of using this inversion strategy to sample the posterior distribution for a linearized cross-borehole tomographic inverse problem conditioned by travel time data.

Many geophysical inverse problems deal with a non-linear forward relation (equation 3), which was also the case for the application of the GDM and PPM (Le Ravalec-Dupin and Noetinger, 2002; Hoffman and Caers, 2004). Moreover, as already discussed, it is of great interest to use complex non-Gaussian prior models that are capable of describing geologically more realistic prior information.

Hansen et al. (2008a) were the first to suggest a way of obtaining realizations from the posterior distribution of non-linear probabilistically formulated inverse problems with both Gaussian and non-Gaussian prior information. In their work, they suggest to use the extended Metropolis algorithm with prior information defined by a geostatistical algorithm. The geostatistical algorithm serves a "black box" prior sampler (Mosegaard, 2006) and the Metropolis rule is used to decide if a proposed realization is rejected or accepted as a realization from the posterior distribution. The examples demonstrated by Hansen et al. (2008a) involved both prior models based on two-point and multiple-point statistics.

2.4 Gradual perturbation strategies

The essential part of the inversion strategies based on GDM and PPM is the perturbation strategies provided by these methods. Likewise, a crucial part of using the extended Metropolis algorithm is to find a way of gradually perturbing the proposed prior realizations in order to make the algorithm perform in an efficient way. The geostatistical prior models used in the work by Hansen et al. (2008a) were based on sequential simulation that was able to perform conditional sequential simulation. Therefore, Hansen et al. (2008a) suggested perturbing the realizations by randomly choosing a subset of the model parameters and re-simulating these model parameters conditionally to the rest of the model parameters. In this way, a new realization that is a perturbation of the previous realization is obtained. The step-length of the perturbation is then controlled by the size of the subset of model parameters that is re-simulated.

A similar re-simulation strategy has later on been suggested by Irving and Singha (2010) and Mariethoz (2010a). In both of these cases, this perturbation strategy is used in conjunction with the extended Metropolis algorithm for inversion of tracer data and hydraulic heads.

Cordua et al. (2010 (paper A12)) demonstrates an example of using the GDM as the perturbation strategy for the extended Metropolis algorithm for a probabilistic formulation of an amplitude vs. off-set reflection seismic inverse problem. An application of PPM as the prior perturbation strategy for the extended Metropolis algorithm might be straight forward, but has not yet been presented. Hansen et al. (2013 (paper A14)) provide a software package with a variety of different prior models that can be used with different perturbation strategies in conjunction with the extended Metropolis algorithm.

In order for a geostatistical sampling algorithm to serve as a valid way of formulating prior information for a probabilistic inverse problem, the output realizations from the algorithm have to be realizations from a certain (not necessarily known) target distribution. One approximate way of verifying if the sampling algorithm converges to and continues sampling a certain target distribution, is to produce a large set of prior realizations and verify (either visually or quantitatively) if the statistical properties (variance, mean, covariance, multiple-point statistics) are preserved in the individual realizations.

2.5 Sequential Gibbs sampling

A joint probability distribution can be expressed as a product of conditional probability distributions by using the product rule

$$\begin{aligned} p(m_N,...,m_1) &= p(m_N \mid m_{N-1},...,m_1)\,p(m_{N-1},...,m_1) \\ &= p(m_N \mid m_{N-1},...,m_1)\,p(m_{N-1} \mid m_{N-2},...,m_1)\,p(m_{N-2},...,m_1) \\ &\;\;\vdots \\ &= \prod_{k=3}^{N} (m_k \mid m_{k-1},...,m_1)...\; p(m_2 \mid m_1)\,p(m_1) \end{aligned} \quad . \qquad (6)$$

This equation is the backbone of sequential simulation. Using this formulation, a realization of the joint probability distribution can be obtained by sequentially simulating one model parameter at a time conditional to the previously simulated model parameters (see e.g. Gómez-Hernández and Cassiraga, 2000).

Having obtained a realization of the joint probability distribution, the Gibbs sampler (Geman and Geman, 1984) can be used to perturb this realization into a new realization by simulating a value from

$$p(m_i \mid m_1, m_2, ..., m_{i-1}, m_{i+1}, ..., m_N) \qquad (7)$$

for one model parameter at a time. However, as already discussed, it is important to be able to control the size of such a perturbation if it has to be used as a prior sampler for an inverse problem. Moreover, perturbing one model parameter at a time may be a very inefficient perturbation strategy.

Hansen et al. (2012 (paper A2)) suggest a Gibbs sampler that in each step simulates values from several model parameters by obtaining a realization of a subset *U* of model parameters $\mathbf{m}_{i\in U}$ from the distribution $p(\mathbf{m}_{i\in U} \mid \mathbf{m}_{i\notin U})$ conditional to the rest of the model parameters not belonging to this subset $\mathbf{m}_{i\notin U}$. A realization of this condition distribution can be obtained through sequential simulation using equation 6. This sampling strategy that combines the Gibbs sampler with sequential simulation is named sequential Gibbs sampling (Hansen et al., 2012 (paper A2)).

Figure 1 below shows an example where the sequential Gibbs sampler is applied. The conditional probabilities used in this example are based on multiple-point statistics obtained from the training image seen in figure 2. A gradual perturbation, as a result of simulating the model parameters related to the gray area, using the sequential Gibbs sampler, is seen between the individual realizations.

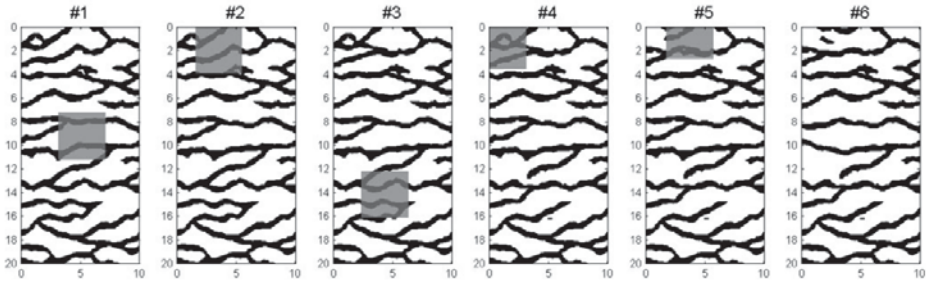


Figure 1. An example of applying the sequential Gibbs sampler. The dark gray field covers the model parameters that are going to be re-simulated, which is seen by the new structures that occur at this location in the next realization (from Hansen et al., 2012 (paper A2)).

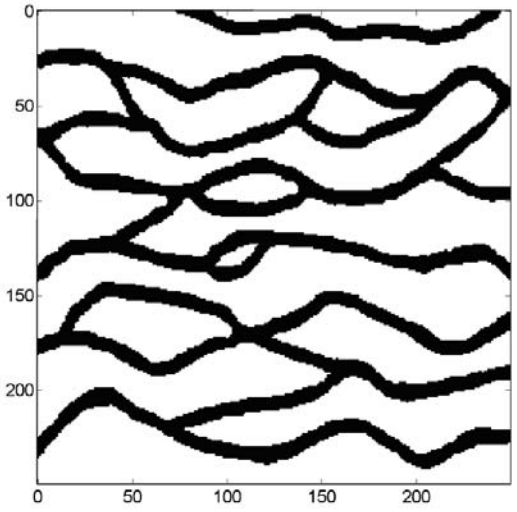


Figure 2. Training image used to obtain multiple-point statistics (from Hansen et al., 2012 (paper A2)).

2.5.1 Detailed balance, aperiodicity, and irreducibility

In order to ensure that the sequential Gibbs sampler will continue sampling the same joint probability distribution over time, the sampling algorithm has to satisfy detailed balance, namely that the probability of a transition from one realization of model parameters $\mathbf{m}_i$ to

another realization $\mathbf{m}_j$ ($\mathbf{m}_i \rightarrow \mathbf{m}_j$) is the same as the probability of a transition $\mathbf{m}_j \rightarrow \mathbf{m}_i$ (Mosegaard and Sambridge, 2002). This will be the case if the number of model parameters that are re-simulated at iteration i is the same as the number of model parameters that are re-simulated in iteration j (see Hansen et al., 2012 (paper A2)). Moreover, the transition probabilities have to remain constant.

Aperiodicity is satisfied if the trivial transition property $\mathbf{m}_j \rightarrow \mathbf{m}_j$ of staying in the same realization is always non-zero. Irreducibility is satisfied if it is always possible to move from any realization $\mathbf{m}_j$ to any other realization $\mathbf{m}_i$ during a sufficient number of re-simulations. If the transition probability $p(\mathbf{m}_j \rightarrow \mathbf{m}_i) = p(\mathbf{m}_i \mid \mathbf{m}_j)$ of moving between two realizations satisfies aperiodicity and irreducibility, in addition to detailed balance, then the joint distribution in equation 6 will be the only equilibrium distribution and the algorithm will converge towards this distribution independent of the starting realization (Mosegaard and Sambridge, 2002).

3 Sampling the solution to inverse problems with geostatistical prior models

Having defined a probabilistic formulation to the inverse problem, a suitable geostatistical prior model, and a perturbation strategy for the prior model, it is now possible to set up an algorithm that samples the posterior distribution of the inverse problem.

3.1 Combining the extended Metropolis algorithm with the sequential Gibbs sampler

Consider a probabilistic inverse problem formulated using equation 4. Moreover, consider that the prior information is described by a geostatistical sampling algorithm that is based on sequential simulation as defined in equation 6. The geostatistical prior model is assumed to sample from an unknown prior probability distribution $\rho_M(\mathbf{m})$. The posterior distribution of such an inverse problem can be sampled by the extended Metropolis algorithm used in conjunction with the sequential Gibbs sampler. This algorithm has the following steps:

1. A starting model is obtained by simulating a full unconditional realization $\mathbf{m}_{start}$ from $\rho_M(\mathbf{m})$ using the geostatistically defined prior model (i.e., an outcome realization from the sampling algorithm). This set of model parameters is becomes a current set of model parameters $\mathbf{m}_{start} = \mathbf{m}_{current}$.

2. A subset *U* of the model parameters $\mathbf{m}_{i\in U}$ from $\mathbf{m}_{current}$ are re-simulated by drawing a realization from the distribution $\rho_M(\mathbf{m}_{i\in U} \mid \mathbf{m}_{i\notin U})$. This leads to a new (proposed) realization, which is also a realization from $\rho_M(\mathbf{m})$.

3. The proposed model is accepted with the probability

$$P_{accept} = \min\left[1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{current})}\right], \tag{8}$$

where $L(\mathbf{m}_{propose})$ and $L(\mathbf{m}_{current})$ are evaluations of the likelihood function for the proposed and current sets of model parameters, respectively. If accepted, the proposed set of model parameters becomes the current model $\mathbf{m}_{current} = \mathbf{m}_{propose}$ and is a realization from the posterior distribution. If not accepted, the proposed set of model parameters is discarded

and the current model is repeated and counts again as a realization from posterior distribution.

4. Step 2 and 3 are repeated in order to obtain multiple realizations that constitute a sample from the posterior distribution.

Note that the size of the subset of model parameters *U* controls the step-length of the sequential Gibbs sampler, which controls the overall computational performance of this algorithm. Using too small step-lengths lead to too many computationally expensive forward calculations (i.e., evaluations of the likelihood function). On the other hand, too large step-lengths lead to too many rejections (i.e., wasted forward calculations).

3.2 Sampling the solution to large scale non-linear inverse problems

Since his seminal work on the full-waveform inverse problem (Tarantola, 1984, 1986, 1988), Albert Tarantola had the vision that geologically realistic prior information for inversion could be learned through a training image of the subsurface (Mosegaard, 2011). Ideally, such a problem should be solved in such a way that not only a single set of model parameters with (near) maximum posterior probability is obtained, but a full characterization (sample) of the posterior distribution should be provided.

In principle, the extended Metropolis algorithm can be used for this. However, if only very little prior information, described by, e.g., an uncorrelated Gaussian distribution, exist, even the best choice of step-length used by the sequential Gibbs sampler will lead to a very inefficient sampling (Hansen et al., 2009). For high-dimensional problems it might even lead to a computationally prohibited inversion, because the probability that the sequential Gibbs sampler will suggest a set of model parameters that is accepted by the Metropolis rule in equation (8) becomes extremely low. In order to compensate for this, a very small step-length can be used, which further contribute to the inefficiency of the sampling.

The effective degree of freedom of the inverse problem is considerably reduced when informative prior information about the model parameters exists. Hansen et al. (2009) showed an example where multiple-point based prior information seemed to be superior with regards to reducing the effective degree of freedom as compared to prior models based on two-point statistics. Such a reduction of the degree of freedom improves the efficiency of the sampling algorithm, which in turn reduces the computational cost of sampling the solution to the inverse problem (Hansen et al., 2009).

Cordua et al. (2012 (paper A4)) demonstrate an example of sampling the posterior distribution of a large scale non-linear tomographic full waveform inverse problem with prior information based on multiple-point statistics. In this study, the prior information was described by multiple-point statistics obtained from a training image. For this purpose, the SNESIM algorithm was used as the geostatistical based prior model. The algorithm presented above, that combines the extended Metropolis algorithm with the sequential Gibbs sampler, was used to sample the posterior distribution of this problem.

The results demonstrate that a sample from the posterior distribution for a high-dimensional (6240 model parameters) full waveform inverse problem with very informative prior information could be obtained in this way. Within reasonable time (45 days on a standard desktop computer with an Intel Core i7 processor) the authors were able to obtain 45 independent realizations from the posterior distribution. Approximately three of these days were spent on the burn-in period, which might have been reduced using a more appropriate optimization procedure during this phase of the algorithm.

In this example, a Gaussian distributed noise component is assumed. The likelihood function takes on the form

$$L(\mathbf{m}) = c \exp\left(-\frac{1}{2}\left(g(\mathbf{m}) - \mathbf{d}^{obs} \right) \mathbf{C}_D^{-1} \left(g(\mathbf{m}) - \mathbf{d}^{obs} \right) \right), \tag{9}$$

where $\mathbf{d}^{obs}$ is a vector of observed data, $\mathbf{C}_D$ is the covariance matrix that describes the (co)variances of the data uncertainties, and *c* is a normalization constant.

The figures 3 and 4 show independent realizations from the prior and posterior distributions, respectively. The multiple-point statistics that is used for the prior information is obtained from the training image seen in figure 2. A comparison between the prior realizations and the posterior realizations reveals the additional information provided by the observed data (and their uncertainties) as compared to the prior information. The synthetic observed data that are used for this inversion have been obtained from the reference model seen in figure 5. A comparison between figure 3, 4, and 5 reveals that the prior information provides the general structural information about the subsurface, and that the observed data provide localized information about where these structures are actually located at this specific site. Moreover, due to the uncertainty related to the observed data, the structures in the posterior realizations are fluctuating around the reference structures seen in figure 5. In this way, an uncertainty estimate (i.e., resolution analysis (Mosegaard, 1998)) of the solution to the inverse problem is obtained. The results do not show any significantly different posterior

realizations, which indicate that multiple considerably different scenarios do not exist for this problem. Alternatively, the algorithm may not yet have located such scenarios.

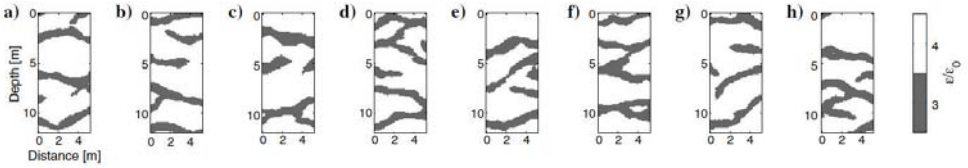


Figure 3. Eight statistically independent realizations from the prior distribution (from Cordua et al., 2012 (paper A4)).

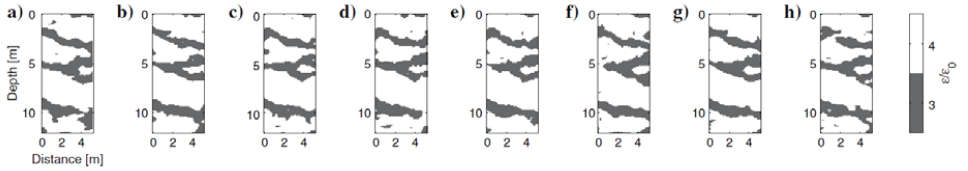


Figure 4. Eight statistically independent realizations from the posterior distribution (from Cordua et al., 2012 (paper A4)).

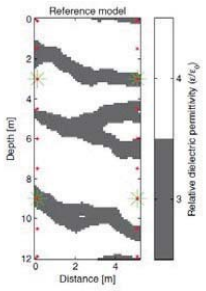


Figure 5. Synthetic reference model. Green asterisks show transmitter positions and the red dots show receiver positions (from Cordua et al., 2012 (paper A4)).

In order to obtain an efficient sampling of the posterior probability distribution, an appropriate degree of perturbation between the prior realizations has to be chosen. Cordua et al. (2012 (paper A4)) suggest a way to determine the degree of perturbation used by the

sequential Gibbs sampler in order to obtain a certain accept probability of the extended Metropolis algorithm.

A sample from the posterior distribution obtained using the extended Metropolis algorithm consists of a set of dependent realizations of the model parameters. All of these realizations can be used for subsequent statistical analysis, such as calculating the mean, variance, or the probability that a certain lithological unit is connected between two points in the subsurface. However, each of these realizations does not contribute with independent information to the statistics. Therefore, the number of statistically independent realizations in the sample should be calculated in order to determine the effective size of the sample. For example, the statistics of 1000 realizations is not worth much if only two statistically independent realizations can be obtained from this sample. In order to calculate this number, autocorrelation between the realizations from the posterior sample can be calculated. In this way, the separation between the posterior realizations that is necessary in order to guarantee that these realizations are statistically independent can be determined (Cordua et al., 2012 (paper A4)).

Figure 6 shows the autocorrelation between a realization from the posterior distribution and its correlation to the next 150000 sets of model parameters obtained from the posterior distribution. These models are not statistically independent because any proposed model in the Metropolis algorithm is a perturbation of a current model or when a proposed model is rejected the current model counts again. Therefore, the autocorrelation analysis of the posterior sample shows some correlation length between successive models. In the example shown in figure 6, statistical independence is obtained after approximately 5800 iterations. This point is approximated as the point at which the autocorrelation curve intercepts the average level of the correlation curve after it is converged to a constant level (marked my a dotted line in figure 6).

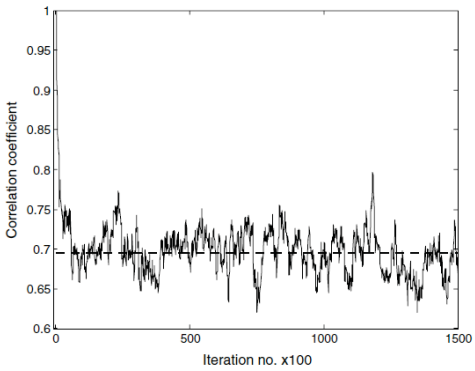


Figure 6. Autocorrelation analysis of the posterior sample, from the full waveform inverse problem, used to determine the number of iterations needed to obtain statistically independent realizations (from Cordua et al., 2012 (paper A4)).

3.3 Algorithm-based prior information

A sample obtained with the extended Metropolis algorithm that uses an algorithm-based prior model (potentially sampled by the sequential Gibbs sampler) has been proven, by Mosegaard and Tarantola (1995) to be consistent with a posterior distribution that is defined as a product between the likelihood function and the unknown prior distribution that is sampled by a prior sampling algorithm.

In the above described example, the prior information is provided by a "black box" geostatistical sampling algorithm, which means that we do not know the prior distribution that is actually sampled by this algorithm. Apart for the input training image (in the case of an algorithm based on multiple-point statistics), such algorithms typically have several input parameters that all influence the output of the algorithm. Therefore, in order to understand the prior information described by a "black box" algorithm, the user of such algorithms should always start out by running the prior sampler with the desired input parameters and/or training image and display these output realizations. For example, this could be done by displaying sequence of prior realizations (i.e., a prior movie) as seen in figure 3. The relative occurrences of different features in the prior movie will reflect the weighting of these features in the prior information. This procedure is necessary because the realizations reflect the prior information about the subsurface, which should reflect the actually prior beliefs about the subsurface. The choice of prior information has enormous impact on the posterior distribution

(i.e., the solution to the inverse problem) and should therefore be chosen wisely and carefully. Even if a mathematical formulation that describe the prior probability distribution exists, the movie strategy should still be applied because it might not be easy to translate the form and parameters of such a formulation into the actual prior information described by such a probability distribution.

As soon as the user of the algorithm agrees that the displayed movie of realizations actually reflects his/hers prior expectations, this prior sampler can be used for the inverse problem and the prior inspection does not have to be performed for inversion of other data if the prior expectations remain the same. Moreover, it should be noted that the choice of a certain perturbation strategy and step-length may influence the outcome of the algorithm-based prior information. Therefore, the realizations used for the movie to be inspected should be based on the desired perturbation strategy and step-length that will be used for the inversion.

If no closed-form mathematical expression exists for the algorithm-based formulation of the prior information, only sampling of the posterior distribution (e.g., through the extended Metropolis algorithm (see Mosegaard and Tarantola, 1995; Mosegaard, 2006)) can be obtained. No optimization for a set of model parameters with maximum posterior probability can be achieved, because in order to do this an evaluation of the posterior probability of any set of model parameters is necessary. This demands that a mathematical expression for both the likelihood function and the prior distribution can be established.

Having obtained these expressions the optimization problem to be solved takes on the following form

$$\arg\max_{\mathbf{m}} \left(\sigma_M(\mathbf{m})\right) = \arg\min_{\mathbf{m}} \left(-\log\left(\rho_M(\mathbf{m})\right) - \log\left(L(\mathbf{m})\right)\right), \tag{10}$$

which is an optimization of equation 4 with respect to the model parameters.

3.4 Improving the pattern reproduction

State-of-the-art geostatistical sampling algorithms strive at producing output realizations with spatial correlations between the model parameters that reflect geological realism. Most of these algorithms use a training image as input and try to produce output realizations that resemble the patterns seen in the input training image. However, this task has proven to be challenging.

Different sampling strategies, such as the direct simulation method (Mariethoz, 2010b) or the cross-correlation simulation strategy (Thamasebi, 2012), have succeeded in obtaining a close match between the input and output patterns of the sampling algorithms.

Cordua et al. (2012 (paper 4)) observed that the patterns seen in the output realizations from the algorithm-based prior model SNESIM do not resemble the patterns seen in the input training image. In particular, when the SNESIM algorithm is used in conjunction with sequential Gibbs sampling, this resemblance was considerably reduced.

In the particular case where the user wishes that the prior information described by the algorithm has high resemblance with patterns seen in the training image, the above observations are not satisfying. It should, however, be noted that this might not always be the desire. All kinds of outputs from such an algorithm-based prior model are satisfying if the output movie of realizations reflects the prior beliefs of the user.

The frequency matching method (Lange et al., 2011 (paper A8)) is a way of quantifying the degree of match between the training image patterns and the patterns in the output realizations. This quantification provides a means of controlling to what degree the patterns in output realizations resembles the patterns seen in the training image. The frequency matching method defines the degree of match between the pattern frequency distribution of a set of model parameters $\mathbf{m}$ (related to the pixels in an image of the subsurface) and the training image TI. This measure is described by a probability distribution $\rho_{FMM}(\mathbf{m}, TI)$ where the best achievable match has the highest probability. $\rho_{FMM}(\mathbf{m}, TI)$ can be defined using different formulations. Examples are the Chi-square dissimilarity measure (Lange et al., 2011 (paper A8)) or the Dirichlet probability distribution (Cordua et al., submitted1 (paper A6)).

Using the Chi-square dissimilarity measure, the frequency matching based prior probability distribution is given as

$$\rho_{FMM}(\mathbf{m}) = \sum_{i \in I} \frac{\left(\pi_i^{TI} - \varepsilon_i^{TI}(\mathbf{m})\right)^2}{\varepsilon_i^{TI}(\mathbf{m})} + \sum_{i \in I} \frac{\left(\pi_i(\mathbf{m}) - \varepsilon_i(\mathbf{m})\right)^2}{\varepsilon_i(\mathbf{m})}, \tag{11}$$

where I is the number of possible pattern combinations given a template and the number of categories used in the training image. π^{TI} is the pattern frequency distribution obtained from the training image and π is the pattern frequency distribution obtained from a set of model

parameters $\mathbf{m}$. ε_i and ε_i^{TI} denote the expected number of counts of the *i*'th pattern obtained from $\mathbf{m}$ and the training image, respectively. These components are defined as

$$\varepsilon_i(\mathbf{m}) = \frac{\pi_i(\mathbf{m}) + \pi_i^{TI}}{n_Z + n_{TI}} n_Z \tag{12}$$

$$\varepsilon_i^{TI}(\mathbf{m}) = \frac{\pi_i(\mathbf{m}) + \pi_i^{TI}}{n_Z + n_{TI}} n_{TI}, \tag{13}$$

where n_Z and n_{TI} are the total number of counts in the pattern frequency distribution related to $\mathbf{m}$ and the training image, respectively.

Using instead the Dirichlet distribution, the frequency matching based prior distribution is given as

$$\rho_{FMM}(\mathbf{m}) = \frac{n_Z!}{\pi_1(\mathbf{m})!,\ldots,\pi_I(\mathbf{m})!} \prod_{i \in I} \left(\frac{\pi_i^{TI} + \pi_i^{\text{prior}}}{n_{TI} + n_{\text{prior}}} \right)^{\pi_i(\mathbf{m})} \tag{14}$$

where π_i^{prior} is the number of counts in *i*'th bin of the prior frequency distribution, which represents the prior expectation of the pattern frequency distribution before the training image histogram is observed. For high values of π_i^{prior} a mismatch between the frequency distributions of a set of model parameters $\mathbf{m}$ and the training image has higher probability than for lower values of π_i^{prior}. For practical implementations of these formulations (equation 11 and 14), see Cordua et al. (submitted1 (paper A6)) and Lange et al. (2012 (paper A9)).

An algorithm that combines the frequency matching method, the sequential Gibbs sampler, and the extended Metropolis algorithm can be used to sample a definition of the prior probability distribution that combines the algorithm-based prior information $\rho_M(\mathbf{m})$ with a prior probability distribution defined by the frequency matching method $\rho_{FMM}(\mathbf{m}, TI)$. The combined prior probability distribution that is sampled by this algorithm is formulated as:

$$\rho_{combined}(\mathbf{m}) = \rho_{FMM}(\mathbf{m}, TI)\rho_M(\mathbf{m}), \tag{15}$$

Using this formulation of the prior probability distribution, the degree of match between the input and output patterns can be controlled and potentially lead to realizations with a high degree of resemblance with the training image patterns (Cordua et al., submitted1 (paper A6)). The algorithm that samples the prior model in equation 15 will be referred to as

the frequency matching sequential Gibbs sampler (FRESEQ sampler). This algorithm has the following steps:

1. A starting model is obtained by simulating a full unconditional realization $\mathbf{m}_{start}$ from $\rho_M(\mathbf{m})$ using the geostatistically defined prior model (i.e., an outcome realization from the sampling algorithm). This set of model parameters is becomes a current set of model parameters $\mathbf{m}_{start} = \mathbf{m}_{current}$.

2. A subset *U* of the model parameters $\mathbf{m}_{i\in U}$ from $\mathbf{m}_{current}$ are re-simulated by drawing a realization from the distribution $\rho_M(\mathbf{m}_{i\in U} \mid \mathbf{m}_{i\notin U})$. This leads to a new (proposed) realization, which is also a realization from $\rho_M(\mathbf{m})$.

3. The proposed model is accepted with the probability

$$P_{accept} = \min\left[1, \frac{\rho_{FMM}(\mathbf{m}_{propose}, TI)}{\rho_{FMM}(\mathbf{m}_{current}, TI)}\right], \tag{16}$$

where $\rho_{FMM}(\mathbf{m}_{propose}, TI)$ and $\rho_{FMM}(\mathbf{m}_{current}, TI)$ are evaluations of the prior probability distribution based on the frequency matching method (see equations 11 and 14) for the proposed and current sets of model parameters, respectively. If accepted, the proposed set of model parameters becomes the current model $\mathbf{m}_{current} = \mathbf{m}_{propose}$ and is a realization from the prior distribution in equation 15. If not accepted, the proposed set of model parameters is discarded and the current set of model parameters $\mathbf{m}_{current}$ is repeated and counts again as a realization from the combined prior distribution.

4. Step 2 and 3 are repeated in order to obtain multiple realizations which constitute a sample from the combined prior distribution $\rho_{combined}(\mathbf{m}) = \rho_{FMM}(\mathbf{m}, TI)\rho_M(\mathbf{m})$.

The combination of the sequential Gibbs sampler, the frequency matching method, and the extended Metropolis algorithm leads to a prior sample strategy that, roughly speaking, rejects prior realizations that have an insufficient match (i.e., fit) with the multiple-point statistics from the training image.

Figure 7 provides a numerical example of using the SNESIM algorithm in conjunction with the sequential Gibbs sampler as prior sampler. The algorithm uses multiple-point statistics from the training image seen in figure 2 as input. Five statistically independent output realizations (from $\rho_M(\mathbf{m})$) are seen in figure 7. An example of using the prior model sampled

by the FRESEQ sampler instead is seen in figure 8, which shows 5 independent realizations from the distribution $\rho_{FMM}(\mathbf{m}, TI)\rho_M(\mathbf{m})$. A comparison between the realizations in figure 7 and figure 8 reveals that FRESEQ sampler provides realizations that have a higher degree of resemblance with the patterns seen in the training image in figure 2 than using the algorithm-based prior model that do not include the frequency matching.

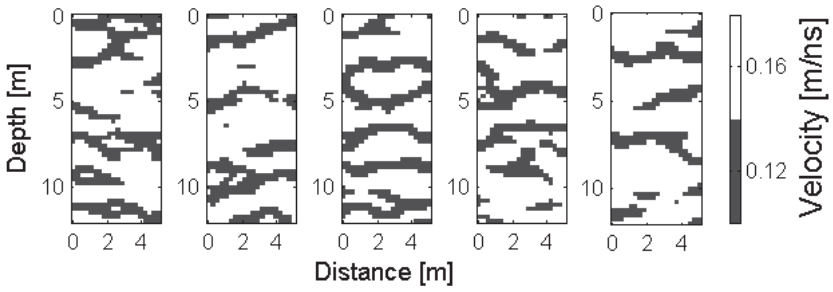


Figure 7. Prior realizations obtained using the SNESIM algoritmh in conjunction with the sequntial Gibbs sampler (from Cordua et al., submitted1 (paper A6)).

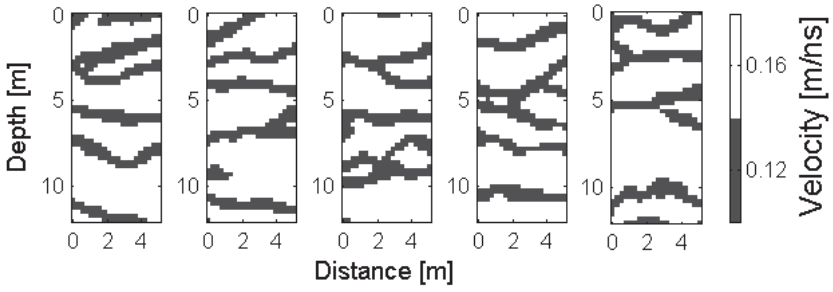


Figure 8. Prior realizations obtained using the FRESEQ sampler (from Cordua et al., submitted (paper A6)).

3.4.1 Sampling of the posterior distribution using a prior model based on the FRESEQ sampler

The FRESEQ sampler can be used as prior sampler for the extended Metropolis algorithm. This leads to an algorithm that uses the Metropolis rule twice, which is also known as using the

Metropolis algorithm in cascade (Mosegaard and Tarantola, 1995). However, cascaded Metropolis has previously only been used in relation to different data sets leading to different likelihood functions. Here, the cascaded Metropolis is instead used for two prior models. This algorithm takes the following steps:

1. A starting model is obtained by simulating a full unconditional realization $\mathbf{m}_{start}$ from $\rho_M(\mathbf{m})$ using the geostatistically defined prior model (i.e., an outcome realization from a sampling algorithm). This set of model parameters becomes a current set of model parameters $\mathbf{m}_{start} = \mathbf{m}_{current}$.

2. A subset *U* of the model parameters $\mathbf{m}_{i\in U}$ from $\mathbf{m}_{current}$ are re-simulated by drawing a realization from the distribution $\rho_M(\mathbf{m}_{i\in U} \mid \mathbf{m}_{i\notin U})$. This leads to a new (proposed) realization, which is also a realization from $\rho_M(\mathbf{m})$.

3. The proposed model is accepted with the probability

$$P_{accept} = \min\left[1, \frac{\rho_{FMM}(\mathbf{m}_{propose}, TI)}{\rho_{FMM}(\mathbf{m}_{current}, TI)}\right], \tag{17}$$

where $\rho_{FMM}(\mathbf{m}_{propose}, TI)$ and $\rho_{FMM}(\mathbf{m}_{current}, TI)$ are evaluations of the frequency matching based prior distribution (see equations 11 and 14) for the proposed and current sets of model parameters, respectively. If accepted, the algorithm proceeds to the next step (step 4). If not accepted, the proposed set of model parameters is discarded and the algorithm jumps back to step 2.

4. Accepted the proposed set of model parameters with the probability

$$P_{accept} = \min\left[1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{current})}\right], \tag{18}$$

where $L(\mathbf{m}_{propose})$ and $L(\mathbf{m}_{current})$ are evaluations of the likelihood function for the proposed and current sets of model parameters, respectively. If accepted, the proposed set of model parameters becomes the current model $\mathbf{m}_{current} = \mathbf{m}_{propose}$ and is a realization from the posterior distribution. If not accepted, the proposed set of model parameters is discarded and the current model is repeated and counts again as a realization from the posterior distribution.

5. Step 2, 3, and 4 are repeated in order to obtain multiple realizations that constitute a sample from the posterior distribution with a prior model defined by equation 15.

Figure 9 shows a synthetic reference model with red curves indicating the transmitter and receiver geometry from where travel times are obtained from this model. Uncorrelated zero-mean Gaussian distributed noise is added to the synthetic data. These travel time data are used as observed data in a probabilistic inversion with a SNESIM based prior using multiple-point statistics from the training image seen in figure 2. The likelihood function that takes into account the Gaussian noise is given in equation 9.

Figure 10 shows realizations from the posterior distribution of this problem, where the SNESIM algorithm is used in conjunction with the sequential Gibbs sampler to describe the prior information. Figure 11 shows realizations from the posterior distribution, where the combined FRESEQ prior sampler is used to describe the prior information. Hence, the prior information used for the posterior realizations in figure 10 and 11 is seen in the prior movies in figures 7 and 8. A comparison between figure 10 and 11 demonstrate that the variability between the individual realizations is reduced when using combined prior model. This confirms that the prior information described by the FRESEQ sampler introduces more information into the posterior distribution than compared to a sample algorithm not based on frequency matching, which leads to an improved resolution (i.e., reduced posterior variability (Mosegaard, 1998)).

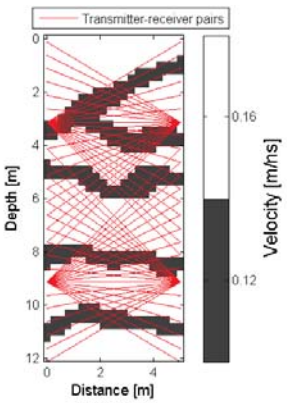


Figure 9. Reference model used to obtain a synthetic data set (from Cordua et al., submitted1 (paper A6)).

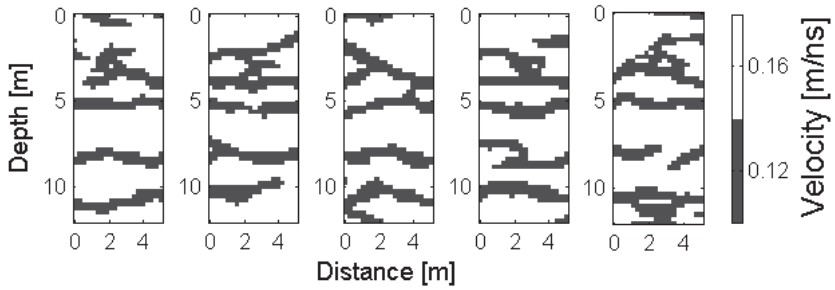


Figure 10. Realizations from the posterior distribution obtained using the SNESIM algorithm in conjunction with the sequential Gibbs sampler as prior sampler (from Cordua et al., submitted1 (paper A6)).

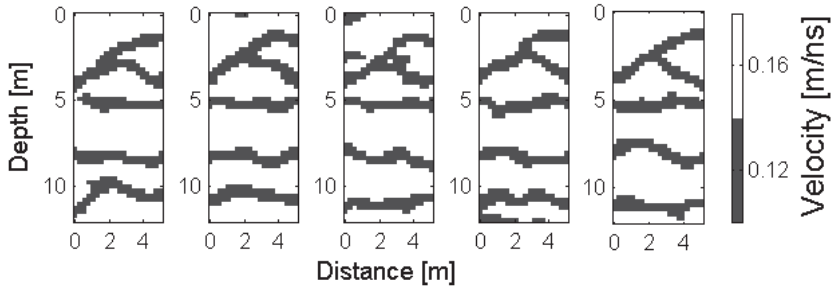


Figure 11. Realizations from the posterior distribution obtained using the FRESEQ sampler as prior sampler (from Cordua et al., submitted1 (paper A6)).

4 From sample models to prior probability distributions

In the previous examples, algorithm-based prior information has been considered. As already discussed, such algorithms produce multiple realizations (i.e., a sample) from some unknown prior probability distribution $\rho_M(\mathbf{m})$ or an approximation to a known distribution is sampled. Examples of such algorithms are SNESIM and SGSIM using a neighborhood.

As discussed above, in order to understand the prior information defined by such algorithms, the user has to inspect a movie of output realizations. However, if we wish to be able to optimize for a set of model parameters with (near) maximum posterior probability, we also need a mathematical expression such that the prior probability of a set of model parameters can be evaluated. Moreover, a closed form mathematical expression allows us to quantify the information content of our prior information, which can be useful in order to quantify the relative influence and information content of different prior models on the posterior distribution (Cordua et al., submitted2 (paper A7)).

In a probabilistic formulation of the inverse problem, the prior information is defined by a prior probability distribution $\rho_M(\mathbf{m})$. This definition of the prior information is typically based on some observations in form of old data sets, previous inversion results, expert knowledge in form of a training image, or borehole information. A set of such observations can be considered as a sample from the prior probability distribution. Therefore, we choose to refer to this type of observations as a sample model.

An example of such a sample model is a training image as seen in figure 12A. In this case, the problem is to establish a (unknown) prior probability distribution that is consistent with the statistics from the training image, such that the training image statistics is a representative outcome from this prior distribution.

One way of attacking this problem is to assume that the unknown prior probability distribution represents a stationary process. This means that the prior probability distribution over a subset of model parameters that are related to a given configuration of pixels (hereafter denoted a “template”) is invariant under a spatial translation of the template. Statistics from the training image is obtained by scanning the training with the template (see, e.g., figure 12B) and the number of different patterns in the training image with this pixel configuration is counted. These counts are stored as a high-dimensional histogram (with dimensions equal to the number of pixels in the template). The concept of extracting a pattern histogram (or frequency distribution) based on two- or multiple-point statistics from a

training image is illustrated in figure 12. Figure 12C shows the non-zero pattern probabilities extracted from the training image in figure 12A using the template in figure 12B.

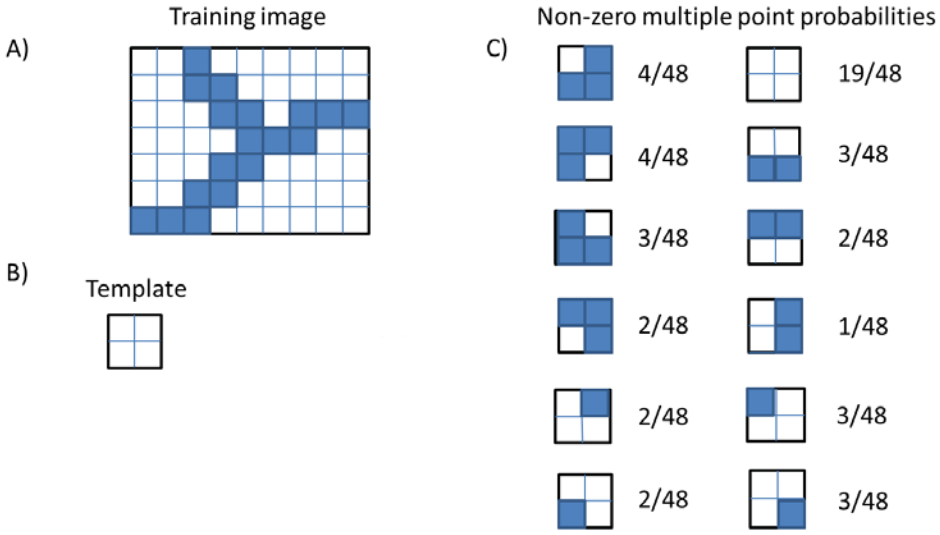


Figure 12. A) Example of a training image. B) Template that defines the pixel configuration used to extract the patterns from the training image. C) Non-zero multiple-point probabilities (patterns that are not shown in this list have zero probability in the training image) (from Cordua et al., submitted2 (paper A7)).

Due to the assumption of stationarity and the assumption that the training image statistics represents a sample from the unknown joint probability distribution, the multiple-point histogram obtained from the training image is an approximation to all marginal probability distributions defined over model parameters in $\rho_M(\mathbf{m})$ related to the template configuration. In other words, these marginal probability distributions (over model parameters related to pixels with a relative configuration defined by the template) will all be the same (due to the assumption of stationarity) and the multiple-point histogram will be an approximation to the marginal probability distribution. That this only leads to an approximation to the marginal distribution is because of the finite size of the sample obtained through the training image. Only an infinitely large sample contains the full information about a probability distribution.

Assume that the pixel values are discrete. Denote by the vector $\mathbf{p}^{\text{joint}}$ the unknown (joint) prior probability distribution, and assume that the vector $\mathbf{p}$ is the known marginal probability distribution obtained from the training image. Since the marginal probability distributions are obtained simply by summations over the probabilities in the joint probability distribution, the problem of determining a joint probability based on a set of its known marginal distributions can be formulated as (Cordua et al., submitted2 (paper A7))

$$\mathbf{A}\mathbf{p}^{\text{joint}} = \mathbf{p}, \tag{19}$$

where the matrix $\mathbf{A}$ describes the summations in the joint distribution that leads to the marginal distributions. Moreover, this system also includes an additional equation that describes that the probabilities of the joint probability distribution have to sum up to one. This problem turns out to be underdetermined (Cordua et al., submitted2 (paper A7)) and additional constraints have to be added in order to reduce the degree of freedom in order to end up with a unique solution to this problem.

An example of adding additional assumptions with the purpose of uniquely determining the joint probability distribution is seen when the joint probability distribution is assumed to be Gaussian. In that case, all two-dimensional marginal distributions from the Gaussian distribution completely describe the joint Gaussian distribution. All the two-dimensional marginal distributions are determined by means, variances, and covariances, which can be obtained through semivariogram analysis (i.e., a pattern histogram based on two-point statistics) of a training image. Hence, in this example the Gaussian assumption provides the additional necessary constraint in order to determine a unique joint probability distribution that is consistent with a set of known two-dimensional marginal probability distribution obtained from training image (two-point) statistics (Cordua et al., submitted2 (paper A7)).

The problem of establishing a probability distribution based on a set of known marginal probability distributions does not only appear in relation to the prior model. A similar problem, which we typically do not pay much attention to, comes up when we establish the data uncertainty model $\rho_D(\mathbf{d})$, which is further used to establish the likelihood function (see equation 4 and 9). In this case, we have a set of observed data and some uncertainty related to these observations described by variances and covariances (even though the covariances are often assumed zero). In order to establish a joint probability distribution (i.e., an uncertainty model) that is consistent with these observations and (co)variances, the distribution is assumed to be Gaussian (because a noise process is typically Gaussian

distributed (e.g., Mosegaard et al. (1997) and Voss et al. (2006)). All two-dimensional marginal probability distributions from the joint Gaussian probability distribution can be determined based on the observations, which represent the mean values of these marginal distributions, and the (co)variances. Hence, in this case the additional constraints needed in order to uniquely determine a joint probability distribution that is consistent with the observations and (co)variances is the Gaussian assumption. With this assumption, all two-dimensional marginal distribution from the joint Gaussian distribution can be determined based on a set of means and (co)variances.

Examples of additional constraints (not necessarily mutually exclusive) that are used in order to establish a unique joint (prior) probability distribution are (cf. Cordua et al., submitted2 (paper A7)): 1) Assuming the joint distribution to be represented by a parametric distribution, such as a Gaussian, 2) assuming stationarity, 3) assuming Markov properties, and 4) assuming maximum entropy.

From the above discussion, it is seen that the problem of establishing a prior probability distribution in case of using two- or multiple-point statistics is highly related. In fact, prior probability distributions based on multiple-point statistics, where correlations between more than two parameters can be considered, are basically extensions from the mathematically more convenient covariance based formulation where only pair-wise correlations are considered.

4.1 Analysis of geostatistical sampling algorithms

Geostatistical sampling algorithms based on sequential simulation typically use a limited set of marginal probability distributions (related to neighborhood or template) obtained either directly from a training image (as depicted in figure 12) or indirectly through a Gaussian assumption with means and (co)variances obtained from a sample model. From these inputs, such sampling algorithms are able to simulate new realizations that are, to some (often unknown) degree, consistent with these marginal distributions (Strebelle, 2002; Mariethoz, 2010b; Deutsch and Journel, 1998).

A general formulation of the joint probability distribution that is sampled by such algorithms can be formulated as (e.g., Whittaker, 1990)

$$\rho_M(\mathbf{m}) = \prod_{i=1}^{N} p(m_i \mid pa(m_i)) \tag{20}$$

where $pa(m_i)$ are the so-called parents to m_i (using a term from the theory of directed graphs), which includes all previously simulated model parameters that m_i depends on (Cordua et al., submitted2 (paper A7)). *N* is the number of model parameters. The marginal probability distributions used to obtain the conditional probabilities p used in equation 20 are either obtained directly or indirectly from a sample model as described above (i.e., a non-parametric or parametric distribution). If m_i always depends on all other model parameters, equation 20 is equal to the formulation used for sequential simulation in equation 6.

The joint probability distribution in equation 20 is also known as a Markov mesh model if the simulation sequence of the model parameters is related to a unilateral path in the image (Kjønsberg et al., 2012; Daly, 2004).

The formulation in equation 20 depends on the simulation sequence (i.e., the order by which the model parameters are simulated). That means that the formulation of the joint probability distribution occurs every time a new simulation sequence is used (Cordua et al., submitted2 (paper A7)). This is not the case for the general formulation of sequential simulation in equation 6 because in this formulation a dependency between all model parameters exists. Moreover, it turns out that the joint probability distribution formulated using equation 20 is not consistent with the known marginal probability distribution obtained from the sample model (Cordua et al., submitted2 (paper A7)). This is not satisfactory with regard to obtaining a consistent prior model.

The joint probability distribution given in equation 20 (based on some set of marginal probability distributions and a certain choice of simulation sequence) is uniquely determined due to the (additional) assumption of Markov properties (i.e., assuming that m_i is only dependent on model parameters related to pixels that are within a certain neighborhood around the pixel related to m_i). However, due to the missing consistency with the known marginal probability distributions this joint probability distribution does not necessarily satisfy equation 19.

The theory of Markov random fields (Castillo et al., 1997) provides a way of describing and sampling a Markovian joint probability distribution that is invariant with respect to the simulation sequence and consistent with the known marginal probabilities. In this formulation, the model parameters from the known marginal probability distributions are associated with nodes in a clique. A clique has the same configuration as the template used to obtain the statistics from the training image. The set of cliques, related to all the known

marginal distributions, will be overlapping (due to the equal overlapping marginal distributions in the stationary joint probability distribution) and form a chain of cliques (see figure 13 for an example). The joint probability distribution defined over such a chain of cliques is given as (Castillo et al., 1997)

$$\rho_M(\mathbf{m}) = \prod_{k=1}^{L} p(r_k \mid s_k), \tag{21}$$

where r_k is a vector of model parameters associated with the *k*'th residual in the chain of cliques and s_k is a vector of model parameters related to the *k*'th separator in the chain of cliques (see figure 13 for an example). *L* is the total number of cliques in the chain of cliques. A simulation of this probability distribution can be done by a simulation that starts in a random clique and proceeds along the chain of cliques (possibly in more than one direction) such that one clique is simulated at a time (Cordua et al., submitted2 (paper A7)). The model parameters within the individual cliques can be simulated using sequential simulation (equation 6). The numbers in figure 13(II) show an example of the order by which the model parameters can be simulated in order to adhere to the chain of cliques. However, other simulation sequences will also satisfy the chain of cliques. The simulation sequence with the individual cliques can be random (e.g. the order by which the nodes marked by ones can be simulated in a random order).

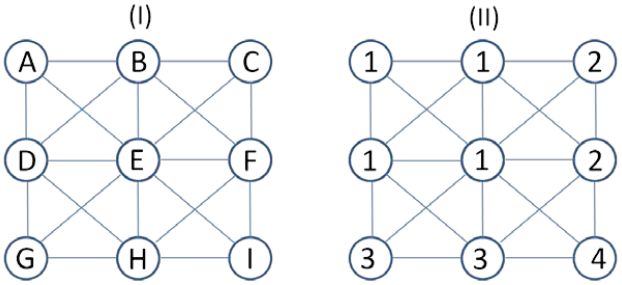


Figure 13. (I) An example of an undirected graph formed by a chain of cliques with the individual cliques defined by the nodes (A,B,D,E), (B,C,E,F), (D,E,G,H), and (E,F,H,I). These cliques are related to a continuous template of 2 x 2 nodes as seen in figure 12. The edges (i.e., connections) in the cliques describe the dependencies between the model parameters related to the nodes. (II) The four residuals related to this chain of cliques are marked by the numbers. Thus, the set of residuals is given as (A,B,D,E), (C,F), (G,H), and (I). The associated separators are given by the sets $\varnothing$, (B,E), (D,E), and (E,F,H) (from Cordua et al., submitted2 (paper A7)).

A realization obtained using equation 21 is seen in figure 14. The marginal probabilities that are used to obtained the conditional probabilities p in equation 21 are obtained by scanning a training image that is a result of using only every third pixel from the training image seen in figure 2 (i.e., a lower resolution image). A continuous template of 3 x 3 pixels (i.e., nodes) is used to obtain the multiple-point statistics from the training image. The realizations are obtained by simulating one clique at a time starting in the upper left corner and simulating one row of cliques at a time (in a raster scan manner). Five realizations obtained in this way are seen in figure 14. The patterns from the training image are, to some degree, recovered. However, a directional "diagonal" effect, which is not seen in the training image in figure 2, is observed in the realizations. This is a well-known effect in realizations obtained from Markov mesh models (Gray et al., 1994), which are very similar to the realizations obtained from equation 21 using this particular simulation sequence. This effect is a result of the order (i.e., direction) by which the model parameters are simulated (Gray et al., 1994).

The directional effects seen in figure 14 are not seen in the results obtained using the SNESIM algorithm as seen in figure 3. Realizations from the SNESIM algorithm are obtained from a distribution given in equation 20 using a random simulation sequence. In this way, no

directional effects are seen. However, using a random simulation sequence while maintaining consistency with the sample model statistics, a conditional probability distribution that describes the dependency to all other model parameters has to be established (see equation 6). In practice, this may not be possible for a non-parametric probability distribution that is based on statistics from a training image, because that number of counts in the statistics decreases when the size of the template increases. Using equation 20, a random simulation sequence can be used within the individual cliques (Cordua et al., submitted2 (paper A7)). Therefore, the directional effects may at least drop with increasing size of the template/clique.

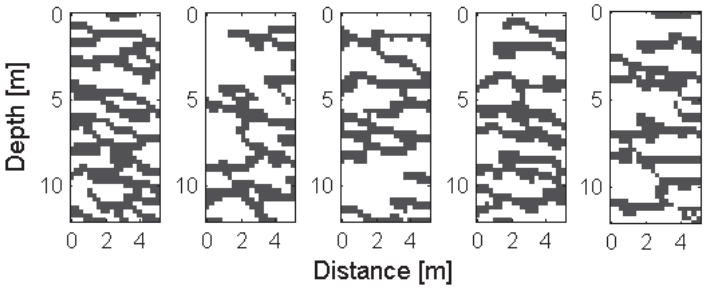


Figure 14. Five realizations obtained from the Markovian joint probability distribution defined in equation 21.

4.2 Accounting for statistical uncertainties and prior expectations of the sample model

As described in the previous sections, a sample model is considered to be based on examples/observations from the field. Similar to other observations, these examples/observations are subject to uncertainty. For example, the marginal probability distribution estimated from the multiple-point statistics obtained from a training image will be an approximation unless an infinitely large training image (i.e., an infinitely large sample) can be provided. Therefore, in order to establish a posterior distribution that is consistent with all information at hand, this statistical uncertainty should be taken into account in the formulation of the prior probability distribution.

This uncertainty can be quantified through the Dirichlet distribution with the observed normalized frequency (i.e., marginal) distribution being the expectation of the distribution and

the variability of the distribution being related to the number of pattern counts obtained from the training image. Hence, the more counts (i.e., the larger a sample) the less uncertainty and the more reliable becomes the estimated marginal distribution.

Moreover, using the Dirichlet distribution as defined in equation 14, a prior frequency distribution π^{prior} can be defined. This distribution represents the prior expectation of the pattern frequency distribution before the frequency distribution of the training image is observed. This means that before the pattern histogram is observed some small frequency (i.e., probability) can be assigned to all possible patterns. Hence, after the pattern frequency distribution has been obtained from the training image, patterns with zero-probability in the training image will still have a (small) non-zero probability in the Dirichlet distribution. Moreover, if the observed pattern frequency (from the training image) is expected to be very uncertain, the prior frequency distribution can be chosen accordingly. High values (i.e., prior counts) in the frequency distribution express an expectation of high uncertainty related to the observed frequency distribution and vice versa.

A manifold that combines the joint (prior) probability distribution (such as for example the formulation in equation 21 above) and the uncertainty distribution of the marginal probability distribution (as described by the Dirichlet distribution) can be expressed as (Cordua et al., submitted2 (paper A7))

$$p(\mathbf{m},\boldsymbol{\pi}) = p(\mathbf{m}\,|\,\boldsymbol{\pi})\,p(\boldsymbol{\pi}). \tag{22}$$

Here, $p(\mathbf{m}\,|\,\boldsymbol{\pi})$ is the joint prior probability distribution given a set of known marginal probability distributions $\boldsymbol{\pi}$. $p(\boldsymbol{\pi})$ represents a probability distribution over the marginal probability distribution that describes the statistical and observational uncertainties related to the marginal probability distributions.

Once the joint manifold over the model parameters and the marginal probability distribution in equation 18 have been defined, the consistent prior probability distribution over the model parameters that takes into account the uncertainty related to the marginal probability distribution is given as the marginal distribution (Cordua et al., submitted2 (paper A7))

$$p(\mathbf{m}) = \sum_{\boldsymbol{\pi}} p(\mathbf{m}\,|\,\boldsymbol{\pi})\,p(\boldsymbol{\pi}) \tag{23}$$

Having obtained this result, it is possible to formulate a mathematical expression of a prior probability distribution that is explicitly formulated and is consistent with the statistics

(and its uncertainties) obtained from a sample model. Realization from this distribution can be obtained using the following algorithm:

1. Draw a realization $\boldsymbol{\pi}$ from the distribution $p(\boldsymbol{\pi})$.

2. Given a realization $\boldsymbol{\pi}$, a realization of the *i*'th model parameter m_i can then be obtained from the distribution $p(m_i \mid m_{j\in J}, \boldsymbol{\pi})$, where $m_{j\in J}$ is the set of previously simulated model parameters. This is an application of sequential simulation.

3. Repeat step 1 and 2 in order to obtain a full realization from equation 23.

All realizations of $\mathbf{m}$ obtained in this way will then be realizations from $p(\mathbf{m})$ as defined in equation 23.

The formulation in equation 23 can be used for sequential simulation such that it can be used for sequential Gibbs sampling. This, in turn, makes it possible to use it in conjunction with the extended Metropolis algorithm for sampling of the posterior distribution. Finally, it can be substituted into equation 10 to be used for optimization with respect to a set of model parameters that have (near) maximum posterior probability.

Having established and/or identified the joint (prior) probability distribution, entropy of the joint probability distribution can be calculated in order to quantify the loss or gain of information between two different joint probability distributions. An example of this is seen in Cordua et al. (submitted2 (paper A7)).

5 Maximum posterior solutions using training image based priors

For computationally very hard problems it may be practically impossible to strive for a sample from the posterior probability distribution. In such cases, one may instead hope to find a solution with maximum or at least high posterior probability. As already mentioned, this can only be achieved if a closed form mathematical expression for both the prior probability distribution and the likelihood function can be evaluated.

When dealing with prior information based on multiple-point statistics obtained from a training image, no mathematical expression for such a prior probability distribution has hitherto existed. This makes it impossible to search for maximum posterior solutions when such prior information is used for the inverse problem. Different approaches have been suggested for a mathematical formulation of such a prior probability distribution.

One way of doing this is to use a formulation such as the Markovian joint distribution as formulated in equation 21 above, which can also include the uncertainty of the marginal probabilities as expressed in equation 23. Another formulation is the frequency matching method (Lange et al., 2012 (paper A9)), which was used in the formulation of the FRESEQ sampler.

The frequency matching method is a probability distribution that has its maximum probabilities in models (i.e., sets of model parameters) with a pattern distribution that has the best possible match with the pattern distribution observed in the sample model (e.g. a training image). The frequency matching method can be formulated using for example a Chi-square dissimilarity measure (Lange et al., 2012 (paper A9)) or a Dirichlet probability distribution (Cordua et al., submitted1 (paper A6)).

In Lange et al. (2012 (paper A9)), the authors use the Chi-square formulation of the frequency matching method to define a prior probability distribution that was used to optimize for a set of model parameters with high posterior probability. This was done for a cross-borehole tomographic travel time inverse problem with prior information based on multiple-point statistics from the training image seen in figure 2. Figure 15 shows the synthetic reference model and one optimized solution for this inverse problem. The synthetic data obtained from the synthetic model are added with uncorrelated zero-mean Gaussian distributed noise. The likelihood function used to formulate the posterior distribution of this problem is given in equation 9. It is clear that this solution honors both the observed data due to the resemblance with the reference model, and the prior information due to the resemblance with the training image.

The prior probability distribution defined by the Chi-square dissimilarity formulation relies on a subjective weighting factor, which is not related to a probabilistic formulation, but found through a regularization strategy (Lange et al., 2012 (paper A9)). Therefore, a purely probabilistic formulation of the frequency matching method based on the Dirichlet distribution was suggested (Cordua et al., submitted1 (paper A6)).

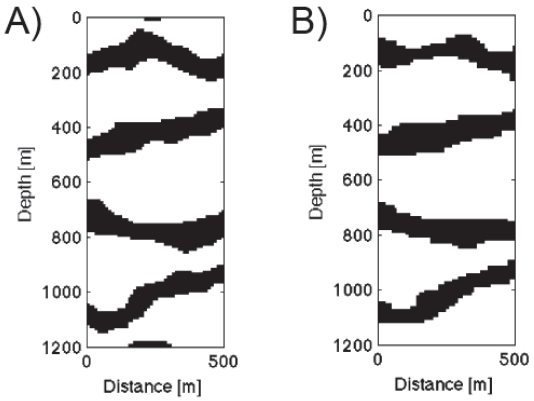


Figure 15. A) Reference model used to obtain travel time data that are used as observed data for the inversion. B) Optimized solution (i.e., set of model parameters) with high posterior probability (from Lange et al., 2012 (paper A9)).

Having established a mathematical expression of the prior probability distribution, simulated annealing can be used to search for the set of model parameters with (near) maximum posterior probability.

For any prior model, algorithm-based or prior models based on a closed form mathematical expression, a movie of output realizations from the prior model should always be used to verify which prior information that is expressed by the prior model. In this way, any arbitrary weighting factor can be calibrated. However, in order to do this in an efficient way, the prior model needs to be formulated in such a way that it is suitable for relatively fast sampling, such as sequential simulation. In this context, the formulation in equation 23 is appealing because it allows performing relatively fast sampling of a formulation of the prior probability distribution that is very similar to the frequency matching method.

5.1 Smooth optimization with the frequency matching method

Most multiple-point based sampling algorithms use a training image that is based on categorical variables. In such a case, the optimization problem for a set of model parameters with maximum posterior probability becomes a combinatorial optimization problem (e.g., as seen in Lange et al., 2012 (paper A9)). Such a problem may be computationally very hard in particular when dealing with high-dimensional inverse problems with computationally intense forward simulators such as full waveform simulation or flow simulation (Cordua et al., 2012 (paper A4); Melnikova et al., 2012 (paper A10)).

An example of getting around this problem is addressed in Melnikova et al. (2012 (paper A10)). Here, a smooth formulation of such a categorical optimization problem in which the model parameters are considered as continuous during the optimization is suggested. The starting model may either consist of categorical or continuous variables. However, due to the prior information based on categorical variables the optimized set of model parameters end up being (almost) categorical. This leads to a shortcut in the optimization, such that the combinatorial optimization is avoided. In this way, the optimization problem can be solved considerably faster than compared to the combinatorial optimization problem.

As it was mentioned in section 2.2, some time is spent on the optimization stage (i.e., burn-in) of the extended Metropolis algorithm. Therefore, an efficient optimization may also save time when searching for a starting model to be used for the subsequent Monte Carlo sampling. For example, in the Monte Carlo sampling of the full waveform inverse problem described earlier, the burn-in period constituted 1/15 of the total computational time. Therefore, it is of interest to obtain a formulation of the prior probability distribution that can be used both for the optimization part and the sampling part, and preferably be used for a smooth optimization. This might be the case for the formulation of the prior probability distribution as described in equation 23.

6 Accounting for imperfect forward models

Typically, the likelihood function (in equation 4) is formulated such that it only takes into account the uncertainties related to the observed data. However, errors related to the forward relation in equations 1 or 3 may also have a significant impact on the solution to the inverse problem. Such errors should also be accounted for in the likelihood function in order to end up with a posterior distribution that is consistent with all the information at hand.

Based on the theory of Tarantola (2005), errors related to the forward relation, referred to as modeling errors, can be taken into account in the likelihood function when this error is assumed Gaussian. Through numerical examples, based on tomographic ground penetrating radar travel time data, Hansen et al. (submitted (paper A11)) demonstrate the effect of such error when using different formulations of the forward relation.

Hansen et al. (submitted (paper A11)) also show that when the modelization error is taken into account the apparent resolution drops, as compared to the case where this error is not taken into account. On the contrary, however, if this error is not accounted for, an erroneously low resolution is obtained, which leads to over optimistic and unrealistic solutions to the inverse problem.

7 Discussion and future research

This section provides a discussion of the research presented so far along with perspectives to future research within this field.

7.1 Mathematical formulations of the prior probability distribution

As we have seen in this thesis, the frequency matching method has been formulated in two different ways. However, recent research in our group has found that it can be demonstrated that the Chi-square dissimilarity measure is an approximation of the Dirichlet distribution. This means that there is a way to relate the weighting factor of the Chi-square dissimilarity formulation of the prior model to a pure probabilistic formulation such that this weighting factor can be formulated using probabilistic reasoning rather than some subjectively chosen weighting factor.

7.2 Optimization of computationally hard inverse problems with multiple-point based prior information

Having obtained a mathematical expression for the prior probability distribution based on multiple-point statistics, it is now possible to use an optimization approach to search for a set of model parameters with high posterior probability. This might be useful either as a fast alternative when sampling of the posterior distribution is computationally prohibited or as a starting point for a sampling algorithm. It would be of interest to test this approach for the tomographic full waveform inverse problem as described in section 3.2.

Tarantola and Valette (1982b) developed an optimization procedure for weakly non-linear inverse problems. This methodology was later on applied to full waveform inverse problems (Tarantola, 1984). This method was based on Gaussian statistics and was a first step away from the regularization based methods where the prior information was wrapped inside a subjectively chosen regularization parameter. The introduction of a mathematical formulation of prior probability distributions based on multiple-point statistics is a further step on the scientific path away from the traditional covariance based approaches towards geologically more realistic prior information. Future research should focus on demonstrating the applicability of such prior model to both optimization and sampling of the solution to large scale inverse problems. It should be remembered that the alternative to not choosing realistic

prior information is some other, potentially smooth Gaussian, prior model, which is known to be geologically unrealistic. Some sort of prior information will always exist in inversion algorithm. Therefore, if you do not choose a prior model, the algorithm will choose for you.

7.3 Quantifying the degree of freedom in an inverse problem

The time needed to obtain a solution to an inverse problem increases with the number of model parameters. For the so-called hard inverse problems, this time increases exponentially with the number of model parameters (Mosegaard, 2006). One way to overcome this is to reduce the effective number of model parameters (i.e., the degree of freedom) by imposing some correlation between the model parameters through the prior information. It can be demonstrated that general algorithms that are not designed for a certain problem will, on average, have the same performance when used for various problems (Mosegaard, 2012). Therefore, prior information about the specific problem turns out to be the only way to improve the performance of the algorithm. Alternatively, an approximation to the inverse problem can instead be solved and the uncertainties related to this approximation can be accounted for in the formulation of the problem (Hansen et al., submitted (paper A11)).

Hansen et al. (2009) demonstrated some examples of how the most informative prior models (i.e., those imposing the highest degree of correlation between the model parameters) resulted in the most computationally efficient sampling of the posterior distribution. The effective degree of freedom was quantified for a Gaussian prior model, which confirmed that the effective degree of freedom was reduced when the degree of autocorrelation among the model parameters was increased. However, in order to do this for prior models based on multiple-point statistics, a mathematical expression for the prior probability distribution is necessary, which has only recently been born. Therefore, this might be a challenge for future research.

7.4 Prior models consistent with a training image

State-of-the-art multiple-point-based geostatistical algorithms strive at reproducing the patterns that are seen in the training image. The FRESEQ prior sampling algorithm that was presented in section 3.4 also aims at reproducing the training image patterns. In section 4 it is described how a prior probability distribution can be established such that it is consistent with the multiple-point statistics from the training image. This approach may not guarantee that

prior realizations from such a distribution resemble the patterns as seen in the training image. On the other hand, such realizations are guaranteed to be consistent with the known multiple-point statistics.

Which form of consistency is the true consistency? In a probabilistic formulation, a formulation that is consistent with the observed statistics from the sample model (e.g., a training image) is the right solution. In this way, no unnecessary additional non-observed information in included. However, an algorithm that strives at reproducing the patterns of the training image is typically a "black box" algorithm, which is also the case for the prior distribution that is sampled by FRESEQ sampler. Additional information may have been built into the geostatistical sampling algorithm (consciously or unconsciously) with the goal of reproducing these patterns. This additional information may lead to a reproduction of the training image patterns, which is fine if this is actually the prior beliefs. On the other hand, such prior information may be too optimistic in that the patterns seen in a single or few image(s) may be too narrow a (prior) view on the possible geological structures at the site under investigation.

7.5 Sampling or optimization?

Figure 16 shows an example of seven realizations (see figure 16a – g) from a posterior distribution related to a linearized travel time based cross-borehole tomographic inverse problem. A Gaussian prior model is used in this case (for more details see Cordua et al. (2009)). The reference model (i.e., set of model parameters) of this synthetic inverse problem is seen to the left in figure 16 (Ref). It is clear that the crispness (variability of the model parameters) in the reference model is, to a high degree, recovered in the posterior realizations.

Figure 16 (Mean) (see right solution in figure 16) shows the solution with maximum posterior probability, which can also be seen as the mean of an infinitely number of posterior realization (in the Gaussian case). It is seen that this set of model parameters provides a much more smooth solution and do not have the same crispness as seen in the reference model and the posterior realizations. It might be computationally convenient to calculate the maximum posterior solution instead of sampling the posterior solution. However, the maximum posterior solution provides a solution that, at least in the Gaussian case, does not honor the prior statistics, and leads to smooth solutions which are unlike the typical realizations from the prior model.

This problem should also be considered when dealing with other prior models. An example is the Markovian joint probability distribution defined in equation 21. The maximum of this probability distribution is the particular case where all the conditional probability distributions in the product all have maximum probability. In such a case, a realization from this probability distribution may not adhere to the multiple-point statistics from a sample model, but be biased towards the most probable patterns. Consequently, a maximum posterior distribution based on such a prior model may lead to overly smooth solutions or solutions that do not honor the prior statistics from the sample model.

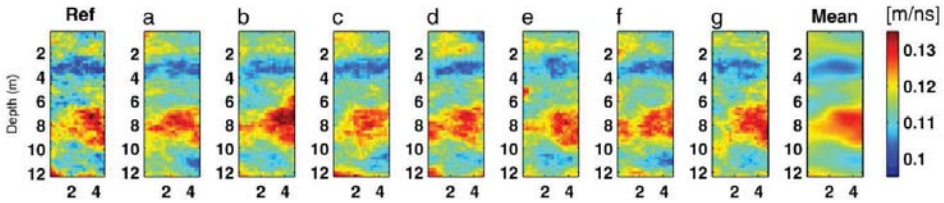


Figure 16. Ref) Reference model used to obtain synthetic travel time data. a – g) Posterior realizations from the posterior distribution. Mean) The mean of infinitely many posterior realizations, which corresponds to the solution with maximum posterior probability (from Cordua et al., 2009).

Instead of searching for a maximum posterior solution it might instead be of interest to search for a most probable posterior realization. In the Gaussian case, the posterior realizations are distributed according to the Chi-square distribution as a function of the distance from the mean of the Gaussian distribution (see section 6.8 in Tarantola (2005)). This means that for high-dimensional distributions (e.g., more than 20 model parameters), the probability of having a realization close to the mean (i.e., near the maximum posterior probability) is extremely low. Hence, the posterior mean solution is in this case an extremely unlikely realization. The highest density of realizations in a Gaussian distribution is located near a sphere with a radius equal to the degree of freedom of the Gaussian distribution. Hence, if a single solution to the inverse problem should be considered, a solution that is located near this sphere should be chosen, which is a solution with most likely statistical properties. Such a solution may be more interesting than the solution with maximum posterior probability, which is an extremely unlikely posterior realization that does not adhere to the prior information (such as the crispness seen in the example in figure 16).

7.6 Stationary and non-stationary prior models

When a training image is assumed to be a realization from a stationary process and at the same time do not contain any directional information, the statistics used to approximate the marginal probability distribution can be obtained not only from the original training image, but also from the image that results from flipping the training image upside down and left right. Hence, in this way four training image can be used to obtain slightly better statistics, even though these images are statistically dependent.

A Markovian joint probability distribution is based on a set of conditional probability distributions (see equation 21). If a prior probability distribution, based on this formulation, has to be non-stationary, the individual conditional probability distributions can no longer be assumed to be equal and an estimate for each of these conditional distributions has to be determined. This is very local information, which might only rarely be accessible.

The training image might not look stationary. However, this can be considered as a scale dependent problem, which means that if a much larger training is provided such a training image could appear stationary. Hence, all training images can ideally be assumed to be a realization from a stationary process. On the other hand, a non-stationary process can in principle not be verified from a single training image.

7.7 Geologically realistic prior models

Outputs from state-of-the-art geostatistical prior models do not yet completely resemble the real world geological structures. Thus, more research needs to be put into this field in order to further approach geological realism. Promising examples within object based geostatistical modeling indicate that this may be possible.

Developments with fractal geometry may be a way of describing geological structures. In particular, this might be useful in order to capture structures at various scales within the same prior model.

However, very detailed and highly informative prior information may rarely be achievable because such information demands local and specific information about the subsurface, which is notoriously unknown since we are dealing with an inverse problem related to an unknown subsurface.

8 Conclusions

Probability theory provides a framework for formulating a solution to inverse problems that are consistent with all conceivable information about the subsurface in form of indirect observed geophysical data and geological prior information. The sources of information are provided in form of probability distributions that need to be established in order to obtain the posterior probability distribution. Given that trustworthy sources of information about the subsurface are provided, the probabilistic formulation of the inverse problem provide an explicit solution to the inverse problem that is consistent with the available information.

Moreover, a combination of probabilistic inverse problem theory and geostatistics provides a means of solving inverse problems with prior information that is founded on geological knowledge, which is a considerably development compared to the early inversion strategies based on subjectively chosen regularization parameters. In addition, the probabilistic formulation provides a means of determining uncertainties and ambiguities in the solution to the inverse problem.

Algorithms exist that are able to obtain a sample from the posterior distribution. This is typically necessary in the case of non-linear inverse problems, non-Gaussian prior information, or if the prior information can only be provided through a sampling algorithm. An example of using the sequential Gibbs sampler in conjunction with the extended Metropolis algorithm for sampling the posterior distribution of a full waveform inverse problem with multiple-point based prior information has been provided. In this case, the prior information was described by a geostatistical sampling algorithm.

Geostatistical sampling algorithms based on multiple-point statistics typically use a training image as input and the output is multiple realizations that to some degree honor the multiple-point statistics from the training image. One approach is to strive for an algorithm that is capable of reproducing the patterns that are seen in the training image. Such an algorithm (i.e., the FRESEQ sampler) that can be used to describe prior information for an inverse problem has been presented.

Another approach to this problem is to establish a prior probability distribution that is consistent with the multiple-point statistics from the sample model such as a training image. In this case, the goal is to obtain prior realizations that are consistent with the training image statistics. However, these realizations do not necessarily reproduce the patterns seen in the training image. An analysis of how to establish such a consistent prior probability distribution has been provided. The findings demonstrate that some assumptions about the probability

distribution have to be made in order to obtain a unique probability distribution that is consistent with the statistics from the training image.

Geostatistical sampling algorithms that are based on sequential simulation have been investigated. It is found that prior probability distributions that are sampled by such algorithms are not necessarily consistent with the training image statistics. However, from Markov random field theory an example of a consistent prior probability distribution is provided. Moreover, an example of an algorithm that is consistent with the observed statistics from the training image and at the same time takes into account the uncertainty related to this statistics has been proposed.

Another example of inconsistency that sneaks into the solution of the inverse problem is through approximate forward relations. Approximations to the forward relation are typically used in order to obtain computationally efficiency in the solution to the inverse problem. The uncertainty related to such approximations can be accounted for through the uncertainty model. In this way, a consistent posterior distribution that leads to a correct and unbiased resolution analysis of the solution to the inverse problem can be obtained.

Different approaches for establishing a closed form mathematical expression for the prior probability distribution for multiple-point based prior information have been suggested. One of these methods is frequency matching method, which has been used to obtain a set of model parameters with (near) maximum posterior probability for a tomographic travel time inverse problem with prior information obtained from a training image.

Populært dansk resumé

Her følger et dansk resumé af PhD afhandlingen med den oversatte titel "Multi-scenarie generering af undergrundsmodeller – Konsistent integration af information fra geofysiske og geologiske data gennem kombination af sandsynlighedsbaseret invers problem teori og geostatistik":

Karakterisering af undergrundens egenskaber og strukturer har stor betydning i forbindelse med 1) vurdering af drikkevandsressourcers beskyttelse, 2) udbredelses- og volumenbestemmelse af olie-, gas- og vandreservoirer, 3) forudsigelser af produktionsdata fra reservoirer og 4) forståelse af geologiske processer og dynamikker. Geovidenskabelige inverse problemer beskæftiger sig med bestemmelsen af undergrundens egenskaber ud fra indirekte observationer af undergrunden. Eksempler på sådanne observationer er seismiske, elektromagnetiske eller tyngdemålinger.

Denne afhandling beskæftiger sig med sådanne inverse problemer. Der bruges her en sandsynlighedsteoretisk formulering af problemet, hvilket giver mulighed for at kombinere de forskellige stumper af information om undergrunden, der er til rådighed. Sådanne informationer kommer fra 1) de fysiske målinger af undergrunden og deres usikkerheder og 2) prior information (dvs. før-viden eller forhåndsviden) om undergrunden. En sådan prior information stammer fra ekspertviden i form af træningsbilleder, som repræsenterer konceptuelle strukturer af undergrunden, eller fra andre målinger i et tilsvarende område. I en sandsynlighedsteoretisk formulering er disse forskellige stumper af information givet som sandsynlighedsfordelinger, der kan kombineres til en posterior information (dvs. efter-viden) i form af en sandsynlighedsfordeling.

Geologisk ekspertviden kan kvantificeres ved hjælp af geostatistiske algoritmer. Disse algoritmer kan generere mange undergrundsmodeller (dvs. et sample), som er konsistente med prior sandsynlighedsfordelinger, der baserer sig på information fra f.eks. træningsbilleder. I denne afhandlingen beskrives det, hvordan prior information beskrevet med geostatistiske algoritmer, kan integreres med sandsynlighedsbaserede inverse problemer. Ligeledes beskrives en algoritme som kan bruges til at generere mange undergrundmodeller, som er konsistente med posterior sandsynlighedsfordelingen for sådan et inverst problem. De mange undergrundsmodeller kan bruges til at beregne usikkerheder og undersøge, om der eksisterer flere scenarier af løsninger, som tilfredsstiller den information der er til rådighed. Samtidig giver integration af ekspertviden løsninger til det inverse problem, som er mere geologisk realistiske. Endeligt viser det sig at prior information om

undergrunden har den ekstra egenskab, at den reducerer det inverse problem sådan at det bliver nemmere at løse.

Geostatistiske algoritmer fungerer som ”sorte bokse”, hvilket skal forstås på den måde at den prior sandsynlighedsfordeling, som samples af disse algoritmer, er ukendt. Kun outputtet fra disse algoritmer giver et indblik i denne fordeling. Ulempen ved en sådan ”sort boks” algoritme er, for det første, at relationen mellem prior sandsynligheden og den tilgængelige ekspertviden er ukendt, hvilket kan føre til inkonsistens mellem den tilgængelige ekspertviden og den prior viden, der er indeholdt i prior sandsynlighedsfordelingen. Denne prior viden påvirker posterior sandsynlighedsfordelingen (dvs. løsning til det inverse problem), som derfor kan risikere at blive ukorrekt. For det andet, kan det inverse problem ikke formuleres som et optimeringsproblem når prior sandsynlighedsfordelingen er ukendt, hvilket kan være et nødvendigt alternativ til sampling i tilfældet af højdimensionale og beregningstunge problemer.

I denne afhandlingen analyseres nogle af disse geostatistiske algoritmer og det beskrives, hvordan en konsistent prior sandsynlighedsfordeling kan etableres. Ekspertviden, i form af f.eks. et træningsbillede, er på samme måde som de indirekte observerede data, behæftede med usikkerheder. En metode, der kan tage højde for denne usikkerhed, er beskrevet.

En alternativ anvendelse af træningsbilleder handler om at opnå et output fra de geostatistiske algoritmer, som har strukturer med høj grad af lighed med de strukturer som ses i træningsbilledet. En algoritme der efterstræber dette og som kan bruges til at beskrive prior information til det inverse problem bliver foreslået.

Afslutningsvis beskrives en metode, der kan tage højde for inkonsistens som relaterer sig til de fysiske modeller som bruges til at kvantificere sammenhængen mellem undergrundens egenskaber og de observerede data. I de tilfælde hvor der anvendes en approksimation til denne model, vil det påvirke løsningen til det inverse problem, medmindre der tages højde for denne fejl. Eksempler på både at tage højde for og ignorere denne fejl bliver beskrevet.

Fælles for de metoder og stratgier, der er beskrevet i denne afhandling er at de alle stræber efter af opnå en løsning til det inverse problem, som er konsistent med den information der er til rådighed, og ikke basere sig på subjektive eller ukendte valg. Fremtidige studier vil, gennem anvendelser, vise de praktiske betydningen af denne fremgangsmåde.

References

Arpat, G.B., 2005, Sequential simulation with patterns. Ph.d. dissertation,

Stanford University. Pp. 184.

Arpat, G.B., J. Caers, 2007, Conditional simulation with patterns. Mathematical

Geology, 39(2), 177 – 203.

Barnes, C., M. Charara and A. Tarantola, 1996, Geological Information and the Inversion of Seismic Data. In: Inverse methods: interdisciplinary elements of methodology, computation, and applications, B.H. Jacobsen, K. Mosegaard, P. Sibani (Eds.), Lecture notes in earth sciences, 63, Springer. Pp. 341.

Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of cross-well seismic data. Geophysical prospecting, 56, 897 – 906.

Barros, L.D., M. Dietrich and B. Valette, 2010, Full waveform inversion of seismic waves reflected in a stratified porous medium. Geophysical Journal International, 182, 1543 – 1556.

Binley, A., G. Cassiani, R. Middleton and P. Winship, 2002, Vadose zone flow model parameterisation using cross-borehole radar and resistivity imaging. Journal of Hydrology, 267, 147 – 159.

Belina, A., J.R. Ernst and K. Holliger, 2009, Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches. Journal of Applied Geophysics, 68(1), 85 – 94.

Boucher, A., P.C. Kyriakidis, C. Cronkite-Ratcliff, 2008, Geostatistical Solutions for Super-Resolution Land Cover Mapping. IEEE Transactions on geoscience and remote sensing, 46 (1), 272 – 283.

Buland, A. and H. Omre, 2003, Bayesian linearized AVO inversion. Geophysics, 68 (1), 185-198.

Buursink, M., T. Johnson, P. Routh and M. Knoll, 2008, Crosshole radar velocity tomography with finite-frequency Fresnel volume sensitivities. Geophysical Journal International, 172, 1 - 17.

Caers, J., 2003, History Matching Under Training-Image-Based Geological Model Constrains, SPE Jounral, 218 – 226.

Caers, J. and T. Hoffman, 2006, The probability perturbation method: A new look at Bayesian inverse modeling. Mathematical Geology, 38 (1), 81 – 100.

Castillo, E., J.M. Gutiérrez, A.S. Hadi, 1997, Expert systems and probabilistic network models. Springer-Verlag New York, Inc. Pp. 605.

Daly, C., 2004, Higher order models using entropy, Markov random fields and sequential simulation. In: O. Leuangthong and C. V. Deutsch (eds.), Geostatistics, Banff 2004, 215-224.

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, no. 1, 71 – 84.

Cordua, K.S., M.C. Looms, and L. Nielsen, 2008, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data. Vadose Zone Journal, 7, 263–271.

Cordua, K.S., T.M. Hansen, and K. Mosegaard, 2010, Nonlinear AVO inversion using geostatistical a priori information. IAMG 2010, Budapest, Hungary, 29 August - 2 September.

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., T.M. Hansen, and K. Mosegaard, submitted1, Improving the pattern reproducibility of multiple-point-based prior models. Mathematical Geosciences.

Cordua, K.S., T.M. Hansen, and K. Mosegaard, submitted2, On sequential simulation of incompletely known random fields. Mathematical Geosciences.

Dahl-Jensen, D., K. Mosegaard, N. Gundestrup, G. Clow, S. Johnsen, A. Hansen, and N. Balling, 1998, Past Temperatures Directly from the Greenland Ice Sheet. Science, 282 (5387), p. 268.

Delfiner, P. and J.P. Chiles, 1977, Conditional simulation, a new Monte Carlo approach to probabilistic evaluation of hydrocarbon in place. Technical Report N-256, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.

Deutsch, C.V. and A.G. Journel, 1998, GSLIB: Geostatistical Software Library, Oxford Press. Pp. 369.

Deutsch, C.V. and L. Wang, 1996, Hierarchical object-based modeling of fluvial reservoirs. Mathematical Geology, 28, 857 - 880.

Ernst, J.R., A.G. Green, H. Maurer and K. Holliger, 2007, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. Geophysics, 72(5), J53–J64.

Gauthier, O., J. Virieux and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. Geophysics, 51, 1387–1403.

Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell., 6, 721–741.

Gómez-Hernández, J., E.F. Cassiraga, 2000, Sequential conditional simulations with linear constrains. In: Monestiez, P., Allard, D., Froideveaux, R. (Eds.), Geostatistics'2000 Cape Town. Geostatistical Association of Southern Africa.

Gray, A.J., J.W. Kay, and D.M. Titterington, 1994, An Empirical Study of the Simulation of Various Models Used for Images. IEEE Transactions on pattern analysis and machine intelligence, 16(5). 507 – 513.

Guardiano, F. and M. Srivastava, 1993, Multivariate geostatistics: beyond bivariate moments. In: Geostatistics - Troia, pp. 133 – 144. Kluwer Academic, Dordrecht.

Hansen, T.M., A.G. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian theory and geostatistics. Geophysics, 71, R101–R111.

Hansen, T.M., K. Mosegaard, and K.S. Cordua, 2008a, Using geostatistics to describe complex a priori information for inverse problems, in J. M. Ortiz, and X. Emery, eds., Geostatistics 2008: Chile, vol. 1.

Hansen, T.M. and K. Mosegaard, 2008b. VISIM: Sequential simulation for linear inverse problems. Computers and Geosciences, 34, 53–76.

Hansen, T.M., K. Mosegaard, and K.S. Cordua, 2009, Reducing complexity of inverse problems using geostatistical priors. IAMG 2009, Stanford, Ca.

Hansen, T.M., K.S. Cordua, B.H. Jacobsen, and K. Mosegaard, accepted for publication, Accounting for imperfect forward models in geophysical inverse problems - exemplified for cross hole tomography. Geophysics.

Hansen, T.M., K.S. Cordua, and K. Mosegaard, 2012, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. Computational Geosciences, 16, 593 – 611.

Hansen, T.M., K.S. Cordua, M.C. Looms, and K. Mosegaard, 2013, SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 1 - Methodology . Computers & Geosciences, 52, 470 – 480.

Hoffman, B. T. and J. Caers, 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.

Honarkhah, M., 2011, Stochastic simulation of patterns using distance-based pattern modeling. Ph.d. dissertation, Stanford University. Pp. 388.

Hu, L.Y., 2000, Gradual deformation and iterative calibration of Gaussian-related stochastic models. Mathematical Geology, 32 (1), 87 – 108.

Hu, L.Y., 2002, Combination of dependent realizations within the gradual deformation method. Mathematical Geology, 34 (8), 953 – 963.

Irving, J.D. and K. Singha, 2010, Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resources Research, 46 (11), W11514.

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

Journel, A.G. and E.H. Isaaks, 1984, Conditional indicator simulation: Application to a Saskatchewan Uranium Deposit, Mathematical Geology, 16 (7), 685 – 718.

Journel, A.G. and F. Alabert, 1989, Non-Gaussian data expansion in the earth science. Terra Nova, v. 1:123 – 134.

Journel, A.G. and T. Zhang, 2006, The necessity of a multiple-point prior model. Mathematical Geology, 38 (5), 591–610.

Kjønsberg, H., M. Stien, O. Kolbjørnsen, B. Fjellvoll, and P. Abrahamsen, 2012, Using Multiple Grids in Markov Mesh Facies Modeling. Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

Khan, A. and K. Mosegaard, 2002, An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data. J. of Geophys. Res, 107 (E6), pp. 19 – 44.

Krige, D.G., 1951. A statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand. Master's thesis, University of Witwatersrand. Pp. 272.

Lange, K., J. Frydendall, K.S. Cordua, T.M. Hansen, Y. Melnikova, and K. Mosegaard, 2012. A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information. Mathematical geosciences, 44, 783 – 803.

Lange, K., K.S. Cordua, J. Frydendall, T.M. Hansen, and K. Mosegaard, 2011, A Frequency Matching Method for Generation of a Priori Sample Models from Training Images. IAMG 2011, Salzburg, September 5- 9.

Le Ravalec-Dupin, M. and B. Noetinger, 2002, Optimization with the gradual deformation method. Mathematical Geology, 34 (2), 125 – 142.

Mariethoz, G., P. Renard, and J. Caers, 2010a, Bayesian inverse problem and optimization with Iterative Spatial Resampling. Water resources research, 46, W11530.

Mariethoz, G., P. Renard, and J. Straubhaar, 2010b, The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46, W11536.

Matheron, G., 1967, Elements pour une Théorie des Milieux Poreux. Masson, Paris.

Matheron, G., H. Beucher, C. de Fouquet, A. Galli, D. Guérillot, and C. Ravanne, 1987, Conditional simulation of the geometry of fluvio-deltaic reservoirs. SPE 16753.

Melnikova, Y., K.S. Cordua, and K. Mosegaard, 2012, History Matching: Towards geologically reasonable models. EAGE conference, Dubai, United Arab Emirates, 25 – 28 November 2012.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, 1953, Equation of state calculations by fast computing machines. J. Chem. Phys., 1 (6), 1087–1092.

Mosegaard, K., 2006, Monte Carlo analysis of inverse problems. Doctoral thesis, faculty of science, University of Copenhagen. Pp. 61.

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse Problems, 14, 405–426.

Mosegaard, K., 2011, Quest for consistency, symmetry, and simplicity - The legacy of Albert Tarantola. Geophysics, 76 (5), W51 – W61.

Mosegaard, K., 2012, Limits to nonlinear inversion. Applied Parallel and Scientific Computing. Proceedings of the PARA 2010 Meeting, Springer, 11 - 21.

Mosegaard, K. and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research, 100, 431–447.

Mosegaard, K. and M. Sambridge, 2002, Monte Carlo analysis of inverse problems. Inverse Problems, 18, R29 – R54.

Mosegaard, K., S. Singh, D. Snyder, H. Wagner, 1997, Monte Carlo analysis of seismic reflections from Moho and the W reflector. Journal of geophysical research, 102(B2), 2969 - 2981.

Mustapha, H. and R. Dimitrakopoulos, 2010, High-order stochastic simulation of complex spatially distributed natural phenomena. Mathematical Geosciences, 42, 457 – 485.

Remy, N., A. Boucher and J. Wu, 2009, Applied geostatistics with SGeMS: A user's guide. Cambridge University Press, Cambridge. Pp. 264.

Rothman, D.H., 1985, Nonlinear inversion, statistical mechanics and residual statics corrections. Geophysics, 50, pp. 2784-2796.

Rothman, D.H., 1986, Automatic estimation of large residual statics

Corrections. Geophysics, 51, 332-346.

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

Suzuki, S. and J. Caers, 2008, A Distance-based Prior Model Parameterization for Constraining Solutions of Spatial Variability. Mathematical Geosciences, 40, 445 – 469.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49, 1259–1266.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data. Geophysics, 51, 1893–1903.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. Pure and Applied Geophysics, 128, 1-2, 365–399.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation. Society of Industrial and Applied Mathematics. Pp. 342.

Tarantola, A., and B. Valette, 1982a, Inverse Problems = Quest for Information. Journal of Geophysics, 50, 159–170.

Tarantola, A., and B. Valette, 1982b, Generalized nonlinear inverse problems solved using the least-squares criterion. Reviews of Geophysics and Space Physics, 20 (2), 219–232.

Tahmasebi, P., A. Hezarkhani and M. Sahimi, 2012, Multiple-point geostatistical modeling based on the cross-correlation functions. Computational Geosciences, 16, 779–797.

Tikhonov, A.N., 1963, Resolution of ill-posed problems and the regularization method (in Russian). Doklady Akademii Nauk SSSR, 151, 501–504.

Voss, P., K. Mosegaard and S. Gregersen, 2006, 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.

Wang, L., 1996, Modeling complex reservoir geometries with multiple-point statistics. Mathematical geology, 28 (7), 895 – 907.

Whittaker J (1990) Graphical Models in Applied Multivariate Statistics. Wiley. Pp. 462.

Wu, J., T. Zhang and A.G. Journel, 2008, Fast filtersim simulation with score-based distance. Mathematical Geosciences, 40, 773–788.

Yang, X., A. Klotzsche, G. Meles, H. Vereecken, and J. van der Kruk, 2013, Improvements in crosshole GPR full-waveform inversion and application on data measured at the Boise Hydrogeophysics Research Site. Journal of applied geophysics.

Zhang, T., P. Switzer and A.G. Journel, 2006, Filter based classification of training image patterns for spatial simulation. Mathematical Geology, 38, 63-80.

Appendices

Scientific contributions found in the appendices

The scientific contributions of this thesis have been submitted to and published in international journals and conference proceedings and can all be found in this appendix. The contributions are as follows:

Main contributions

The papers in appendix A1 and A2 introduce the Sequential Gibbs sampler. This sampler makes it possible to sample, in an efficient way, the posterior distribution of a probabilistically formulated inverse problem that combines information from geophysical data with geological (prior) information described by a geostatistical algorithm.

Appendix A3 and A4 provide examples of using the sequential Gibbs sampler in conjunction with the extended Metropolis algorithm to sample the posterior distribution of a computationally hard tomographic full waveform inverse problem.

The papers found in appendix A5 and A6 suggest a methodology that improves the degree of resemblance between the patterns in the input training image and the output realizations of multiple-point geostatistical algorithms that are based on sequential simulation. In particular, this improvement is significant when the prior is sampled by the sequential Gibbs sampler, which, in turn, improves the information content (i.e., resolution) of the posterior distribution.

Appendix A7 is a review and an analysis of existing geostatistical algorithms that are based on sequential simulation. The paper investigates how that probability distribution, which is sampled by such algorithms, is established through marginal probability distributions obtained from a training image. Moreover, it is discussed and investigated how additional assumptions about the distribution are necessary in order to uniquely determine the distribution to be sampled. It is found that typical algorithms based on sequential simulation are not consistent with the marginal distributions obtained from the training image. An alternative formulation for a consistent algorithm is described.

Other contributions

Appendix A8, A9, and A10 concern the development of the frequency matching method, which leads to a closed form mathematical expression of a prior probability distribution based on multiple-point statistics from a training image. The method is further developed such that categorical variables are treated as continuous variables during the inversion, which makes this optimization procedure perform much more efficient. The frequency matching method is tested on a tomographic travel time inverse problem and a history matching (inverse) problem.

The paper in appendix A11 addresses the problem of taking into account the uncertainty related to the forward problem in an inverse problem. Examples are provided that demonstrate how different forward calculations lead to different degrees of uncertainty. Numerical examples show how the uncertainties related to the forward calculations, if ignored, make artifacts appear as well resolved features.

The papers in appendix A12 and A13 are examples of sampling the posterior distributions of an amplitude-versus-offset (AVO) reflection seismic inverse problem.

A software page, named “Sampling the solution to an Inverse Problem with complex Prior Information” (SIPPI), is documented and applied in appendix A14 and A15. SIPPI provides an implementation of the Sequential Gibbs sampler for a variety of prior models. This implementation can be used together with the extended Metropolis algorithm or the rejection sampler to obtain a sample from the posterior distribution of inverse problems with different complex geostatistical based prior information.

Appendix A1:
Sampling informative/complex a priori probability distributions using Gibbs sampling assisted by sequential simulation

Authors:

Thomas Mejer Hansen, Klaus Mosegaard, and Knud Skou Cordua

Published in:

Proceedings of the 14th Annual Conference of the International Association
for Mathematical Geoscience (IAMG 2010)
Budapest, Hungary
29 August - 2 September 2010

Sampling informative/complex a priori probability distributions using Gibbs sampling assisted by sequential simulation

Thomas Mejer Hansen, Klaus Mosegaard, and Knud Skou Cordua[1]

[1]Center for Energy Resources Engineering, Department of Informatics and Mathematical Modelling, DTU, Richard Petersens Plads, Building 321, 2800 Lyngby, Demmark. E-mail: tmeha@imm.dtu.dk

Abstract

Markov chain Monte Carlo methods such as the Gibbs sampler and the Metropolis algorithm can be used to sample the solutions to non-linear inverse problems. In principle these methods allow incorporation of arbitrarily complex a priori information, but in practice current methods allow only relatively simple priors to be used. We demonstrate how sequential simulation can be seen as an application of the Gibbs sampler, and how such a Gibbs sampler assisted by sequential simulation can be used to perform a random walk that generates realizations of a relatively complex random function. We propose to combine this algorithm with the Metropolis algorithm in order to obtain an efficient method for sampling posterior probability densities for non-linear inverse problems.

Keywords: Monte Carlo, prior, sequential simulation, inverse problem.

1. INTRODUCTION

Consider a typical forward problem

$$d = g(m) \quad (1)$$

where a function *g* relates a subsurface model *m* to observational data *d*. Inverse problem theory deals with the problem of inferring properties of *m* from a specific dataset *d*, using eqn. (1) and some prior information on *m*. Tarantola (2005) and Mosegaard (2006) formulated a probabilistic approach to solving inverse problems where a priori information is described by an a priori probability density function (pdf), $\rho_M(m)$, and the data fit associated to a given model is given by a likelihood function, $L(m)$. The solution to such an inverse problem is the a posteriori probability density, which is proportional to the product of the prior and the likelihood:

(2) $$\sigma_M(m) = k\,\rho_M(m)\,L(m)\ ,$$

where *k* is a normalization factor. Hansen et al. (2006) propose an efficient, non-iterative approach using sequential simulation to generate samples of the a posteriori pdf in case *g* is a linear function and both $\rho_M(m)$ and $L(m)$ can be described by Gaussian statistics. Hansen and Mosegaard (2008) relax the Gaussian assumption of this approach, which allows for non-Gaussian a priori distributions.
In practice, *g* is often a nonlinear operator, and $L(m)$ and $\rho_M(m)$ are non-Gaussian, which results in an non-Gaussian a posteriori pdf. Mosegaard and Sambridge (2002) summarize and discuss a number of Monte Carlo based methods for sampling the solution to such problems. Among these we find the rejection sampler, the Gibbs sampler, and the Metropolis algorithm. Each of these methods is guaranteed to sample the a posteriori pdf asymptotically, although the computational efficiency may differ significantly.
They allow an arbitrarily complex noise model and arbitrarily complex a priori information to be used, but they differ in the way the content of the a priori model is presented to the algorithm. A short description of each of these methods, and their demands on the a priori model, is given here.

Rejection sampler

The rejection sampler is perhaps the simplest method for sampling the posterior probability density function, $\sigma_M(m)$ (eqn. (2)). It allows inclusion of complex a priori information, and any black box that generates independent realizations from the a priori probability density function can be used. Rejection sampling works by filtering a list of independent realizations of the a priori model. Each proposed model is accepted with probability

(3) $$p_{acc}(m_{propose}) = L(m_{propose})/M$$

where *M* is larger than (or equal to) the maximum likelihood of all the proposed models. In many cases the maximum likelihood is not known, and one must set *M* to a large value. For large-dimensional problems this typically causes the acceptance probability p_{acc} to be very small, and hence the algorithm to be very inefficient.

Metropolis algorithm

The Metropolis algorithm is a Monte Carlo sampling method based on Markov chains (Metropolis & Ulam, 1949). Mosegaard and Tarantola (1995) describe a generalized Metropolis algorithm that allows analysis of non-linear inverse problems with complex a priori information. The prior information must be quantified in such a way that samples of the a priori probability density, which are (often) a perturbation of the previous sample, can be obtained. Furthermore, one must be able to control the exploratory nature of the prior sampler (i.e. the step length between two successive prior samples) in order to control the efficiency of the algorithm.
Each iteration of the Metropolis algorithm, starting in model m_n consists of two stages: a) exploration and b) exploitation. In the exploration stage, one step of a random walk, sampling the prior, is performed. In other words, an unconditional realization m_{n+1} of the a priori pdf $\rho_M(m)$ in the vicinity of m_n is generated. This is followed by the exploitation stage where the

likelihood of the proposed model is evaluated. If $L(m_n)$ and $L(m_{n+1})$ is the likelihood of m_n and m_{n+1} , respectively, m_{n+1} is accepted with probability:

$$(4) \qquad P_{accept} = \begin{cases} 1 & if\ L(m_{n+1}) > L(m_n) \\ L(m_{n+1})/L(m_n) & otherwise \end{cases}$$

If m_{n+1} is accepted, m_{n+1} becomes the current model. Otherwise m_n remains the current model. Performed iteratively this algorithm will sample the a posteriori pdf, in the sense that output models occur with a frequency proportional to their a posteriori probability.
The computational efficiency of the Metropolis algorithm depends on the structure of the posterior distribution to be sampled. In addition, it is strongly dependent on the exploratory nature (i.e. step length) of the prior sampler. No theoretically correct step length can be found, but Gelman et al. (1996) suggest that a step length giving rise to an acceptance ratio of the Metropolis sampler of about 25-50% is reasonable. In any case, one must be able to adjust the step length in order to successfully apply a specific method to sample the prior.

Gibbs sampler

In each step of the Gibbs sampler (Geman and Geman, 1984), a model parameter m_i is selected at random. Then the conditional probability distribution of m_i , given that the rest of the model parameters are held constant, is computed. Finally, a realization of m_i is drawn from the conditional distribution.
An important property of the Gibbs sampler is that no models are rejected, as is the case for the Metropolis sampler. In most implementations, the main computational task of applying the Gibbs sampler is to compute and draw realization from the conditional pdf. However, it is important to note that computing the local conditional distribution is not a requirement. Any method that is able to generate realizations from the conditional distribution will suffice.
A number of conditions must be satisfied in order to ensure that the Metropolis algorithm and the Gibbs sampler sample the desired distribution. First, *aperiodicity* and *irreducibility* must be satisfied to ensure that the algorithm has a unique *equilibrium distribution* (Mosegaard and Sambridge, 2002). In addition, the desired distribution must be the equilibrium distribution of the algorithm. This is satisfied in a simple way if each pair of neighboring sample points are in *detailed balance*: The probability that a jump takes place from model m_k, to m_l must be equal to the probability that a jump takes place from model m_l to m_k :

$$(5) \qquad P(m_k \rightarrow m_l \mid m_k)\ \rho_M(m_k) = P(m_l \rightarrow m_k \mid m_l)\ \rho_M(m_l)$$

In this manuscript we will focus on ways to quantify complex a priori information such that it can be used with the rejection sampler, the Metropolis algorithm, and the Gibbs sampler. We will demonstrate how geostatistical algorithms, based on sequential simulation, are capable of simulating geological reasonable structures and therefore are suitable for quantifying complex a prior information. These algorithms can be used directly with the rejection sampler. We will show that the Gibbs sampler and sequential simulation are closely related. Specifically we will demonstrate an application of the Gibbs sampler that will enable generating realizations from any stochastic model that can be simulated using sequential simulation. Further we will demonstrate that this sampling algorithm honors detailed balance, such that it will actually sample the stochastic model intended. We suggest using this method, which we refer to as

sequential Gibbs sampling, to control the exploratory nature (i.e. step length) of the Metropolis algorithm. This strategy not only leads to the formulation and solution of inverse problems with complex a priori information, but can also have dramatic effect on the computational efficiency of the Metropolis sampling algorithm.

2 QUANTIFYING PRIOR INFORMATION USING GEOSTATISTICS

Generally speaking geostatistics is an application of random functions to describe spatial phenomena, typically in form of spatial variability in earth models. Geostatistical simulation algorithms have been developed to efficiently generate realization of a number of random function models. Geostatistical simulation algorithms can be divided into two groups, where the underlying random function model is based on 2-point and multiple-point statistics, (Guardiano and Srivastava,1993, Strebelle, 2002). 2-point based geostatistical algorithms take into account spatial variability between sets of two data locations. In case the distribution of the model parameters is Gaussian, one can completely define the underlying random function model using a Gaussian pdf:

$$\rho_M(m) = c\ \exp(-0.5(m - m_{prior})^t\, C^{-1}_{M_{prior}}\,(m - m_{prior})), \quad (6)$$

where m_{prior} is the a priori mean and $C^{-1}_{M_{prior}}$ is the a priori covariance model.

Multiple-point-based geostatistical models have no parametric description. Instead multiple point statistics is inferred from training images. The methodology was initially proposed by Guardiano and Srivastava (1993). Strebelle (2002) developed the first computationally feasible algorithm for categorical training images. Zhang et al. (2006) and Wu et al. (2008) suggested another multiple point based algorithm where patterns from a continuous or categorical training image are used to generate stochastic realizations with features from the training image. Using these techniques one can generate realization of random function models that reproduce geologically realistic spatial variability. State of the art implementations of these algorithms are available through SGeMS (Remy et al., 2008).

2.1 Sequential simulation

Consider N points $\mathbf{u}_1$, $\mathbf{u}_2$, … $\mathbf{u}_N$ and a random field $Z(\mathbf{u})$ describing the interdependence between values of a physical or geological property measured at points $\mathbf{u}_i$. Then, one realization $\boldsymbol{z}(\mathbf{u})$ of the vector

$$\mathbf{Z}(\mathbf{u}) = (Z(\mathbf{u}_1), Z(\mathbf{u}_2), \ldots, Z(\mathbf{u}_n), Z(\mathbf{u}_{n+1}), \ldots Z(\mathbf{u}_N)) \quad (7)$$

can be simulated using sequential simulation as follows: In step i, visit location $\mathbf{u}_i$ and draw a random value $z_i = \boldsymbol{z}(\mathbf{u}_i)$ from $\boldsymbol{Z}(\mathbf{u}_i)$, using the conditional probability density function

$$f_Z(z_i \,|z_1 \ldots z_{i-1}), \quad (8)$$

where $f_Z(z_1 \dots z_N)$ is the joint probability density of the components of **Z**(**u**). That the above procedure will actually sample $f_Z(z_1 \dots z_N)$ follows from the identity $f(s|t)f(t) = f(s,t)$, which (in the general multivariate case) yields

$$(9) \qquad \begin{aligned} f_Z(z_{n+1} \dots z_N \,|z_1 \dots z_n) \quad = \quad & f_Z(z_{n+1} \,|z_1 \dots z_n) \\ & f_Z(z_{n+2} \,|z_1 \dots z_{n+1}) \\ & \vdots \\ & f_Z(z_N \,|z_1 \dots z_{N-1}) \end{aligned}$$

When all locations have been visited once, a realization of **Z**(**u**) 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. Considerable effort has been made in the geostatistical community to efficiently compute conditional probability density functions based on the 2-point and multiple-point stochastic models as described above.

2.2 Sequential Gibbs sampling

Consider a known realization **z**(**u**) of the random function **Z**(**u**) obtained using sequential simulation. If we now, at random, select a model parameter, $z_i = \boldsymbol{z}(\mathbf{u}_i)$, compute the local conditional pdf

$$(10) \qquad f_Z(z_i \,|z_1, z_2, \dots, z_{i-1}, z_{i+1}, \dots, z_n) \ ,$$

and draw a value from it, we get a new realization of the random field **Z**(**u**). If this is repeated iteratively, this will be an application of the Gibbs sampler (Geman and Geman, 1984). The cost of using the Gibbs sampler is that one must be able to generate a realization of the local conditional pdf, which can be done very effectively using methods developed for sequential simulation. We refer to this combination of sequential simulation and Gibbs sampling as sequential Gibbs sampling.

The sequential Gibbs sampler can be used as a prior sampler for the generalized Metropolis algorithm (Mosegaard and Tarantola, 2005). However, in order to control the computational efficiency of this algorithm, some flexibility of the amount of perturbation (i.e. some control of the 'step-length' of the prior sampler) is needed. We suggest considering not just one model parameter at each step of the Gibbs sampler, but a subset *U* of model parameters. Assuming that the model parameters we wish to update belong to such a subset, we need to generate a realization of the conditional pdf

$$(11) \qquad f_Z(z_{i \in U} \mid z_{i \notin U})$$

Recall that we do not need to explicitly calculate the complete conditional distribution in eqn. (11), but only to generate a realization from it. In order to do this we can make use of the sequential simulation approach of eqn. (9), which involves computing only the conditional probability density function for each model parameter in *U* in random order.

In this way we have designed a Gibbs sampler that efficiently samples a prior pdf by utilizing the technique of sequential simulation to draw values from the local conditional pdf. At the

same time this technique provides a means of controlling the step length of the prior sampling by varying the number of model parameters in the subset

Detailed Balance
Mosegaard and Sambridge (2002) demonstrate that the Gibbs sampler satisfies detailed balance. As the Sequential Gibbs sampler is an application of the Gibbs sampler, this property is also ensured for the sequential Gibbs sampler. This means that the random walk performed by running the sequential Gibbs sampler will asymptotically sample the same random function as a traditional sequential simulation.

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 region (i.e. subset) in the physical space and regard all model parameters associated to this area 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 eqn. (11).
3. Use the new model as the starting model and go to 1.

Such an algorithm was proposed by Hansen et al. (2008). They did, however, not make the link to the Gibbs sampler, and they provided no proof that the resulting algorithm samples an equilibrium distribution, nor that such an equilibrium distribution would in fact be the requested a priori model.
As already mentioned, the size of the region of model parameters that is regarded as unknown (in step 1 in the sequence above) can be used to control the step length of the sequential Gibbs sampler. By regarding only a single model parameter as unknown results in a model m_{i+1} which is highly related to the original model m_i . On the contrary, by regarding all model parameters as unknowns leads to m_{i+1} which is statistically independent of m_i .

2.3 Gradual deformation

Techniques that allow a gradual deformation between two realizations of random function have been developed for both 2-point based Gaussian random function models (the Gradual Deformation Method(GDM) by Hu, 2000; Le Ravalec et al., 2000) and multiple-point based random function models (the Probability Perturbation Method (PPM) by Caers and Hoffman, 2006). The use of Gradual Deformation has mainly been as part of an optimization algorithm for data calibration, where iterative gradual deformation have been used to gradually change a starting model until the forward response from the model matches observed data to some satisfactory degree (Caers and Hoffman, 2006). It has been suggested that running such an application several times, generating a set of models that all fit the data, can be used to describe posterior uncertainty. This is however not the case. The resulting model variability of such an approach reflects the choice of optimization algorithm and the level chosen for

acceptable data fit. An effort to address this issue was made by Le Ravalec-Dupin and Noetinger (2002).
If GDM and PPM honors microscopic reversibility they could, however, be used to sample the prior.

3 CONCLUSIONS

We have demonstrated that the Gibbs sampler and the method of sequential simulation are closely related. We have proposed an efficient sampling algorithm - sequential Gibbs sampling combining these two methods. Sequential Gibbs sampling can sample random function models based on 2-point as well as multiple-point based statistics. We have shown that this algorithm satisfies microscopic reversibility and that the equilibrium distribution is in fact the prior probability density. The sequential Gibbs sampler can be run with arbitrary 'step lengths'. The longest step length (where all parameters are updated in one step) result in a new realization from the a prior distribution which in statistically independent to the previous model. A step length of zero returns the same realization as the previous model. This makes the sequential Gibbs sampler well suited as a method for sampling a complex prior in connection with the Metropolis sampler.
Utilization of the sequential Gibbs sampler will allow relatively efficient analysis of the solution to nonlinear inverse problems with complex a priori information.

ACKNOWLEDGEMENT

The present work was sponsored by DONG Energy as part of the research project 'Risk assessment and multiple scenario generation from seismic and geological data'.

REFERENCES

Caers, J. & Hoffman, T., 2006. The probability perturbation method: A new look at bayesian inverse modeling, *Mathematical Geology*, **38**(1), 81 – 100.
Gelman, A., Roberts, G., & Gilks, W., 1996. Efficient metropolis jumping hules, in *Bayesian Statistics 5,*, pp. 599–608, Clarendon press, Oxford.
Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell*, **6**, 721–741.
Guardiano, F. & Srivastava, R., 1993. Multivariate geostatistics: beyond bivariate moments, *Geostatistics-Troia*, **1**, 133–144.
Hansen, T. & Mosegaard, K., 2008. VISIM: Sequential simulation for linear inverse problems, *Computers and Geosciences*, **34**(1), 53–76.
Hansen, T. M., Journel, A. G., Tarantola, A., & Mosegaard, K., 2006. Linear inverse Gaussian theory and geostatistics, *Geophysics*, **71**(6), R101–R111.
Hansen, T. M., Mosegaard, K., & Cordua, K. C., 2008. Using geostatistics to describe complex a priori information for inverse problems, in *VIII International Geostatistics Congress*, vol. 1, pp. 329–338, Mining Engineering Department, University of Chile.
Hu, L. Y., 2000. Gradual deformation and iterative calibration of gaussian-related stochastic models, *Mathematical Geology*, **32**(1), 87–108.
Metropolis, N. & Ulam, S., 1949. The Monte Carlo Method, *Journal of the American Statistical Association*, **44**(247), 335–341.

Mosegaard, K., 2006. *Monte Carlo Analysis of Inverse Problems*, University of Copenhagen.
Mosegaard, K. & Sambridge, M., 2002. Monte carlo analysis of inverse problems, *Inverse Problems*, **18**(3), 29–54.
Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research*, **100**(B7), 12431–12447.
Ravalec, M. L., Noetinger, B., & Hu, L. Y., 2000. The fft moving average (fft-ma) generator: An efficient numerical method for generating and conditioning gaussian simulations, *Mathematical Geology*, **32**(6), 701–723.
Remy, N., Boucher, A., & Wu, J., 2008. *Applied Geostatistics with SGeMS: A User's Guide*, Cambridge Univ Pr.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol*, **34**(1), 1–20.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
Wu, J., Boucher, A., & Zhang, T., 2008. A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM, *Computers and Geosciences*, **34**(12), 1863–1876.
Zhang, T., Switzer, P., & Journel, A., 2006. Filter-based classification of training image patterns for spatial simulation, *Mathematical Geology*, **38**(1), 63–80.

Appendix A2:

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

Authors:

Thomas Mejer Hansen, Knud Skou Cordua, and Klaus Mosegaard

Published in:

Computational Geosciences

Comput Geosci (2012) 16:593–611
DOI 10.1007/s10596-011-9271-1

ORIGINAL PAPER

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

Thomas Mejer Hansen · Knud Skou Cordua · Klaus Mosegaard

Received: 10 May 2011 / Accepted: 12 December 2011 / Published online: 4 January 2012

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 but 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.

Keywords Inverse problem theory · Geostatistics · Geology · Prior information

T. M. Hansen (✉) · K. S. Cordua · K. Mosegaard
Center for Energy Resources Engineering,
DTU Informatics, Technical University of Denmark,
Lyngby, Denmark
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 Eq. 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, $\rho_M(\mathbf{m})$, 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:

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

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], proposes 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 has been limited [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 in-consistent 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 [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 $\mathscr{M}$ and an algorithm A (a *proposal generator*) that is able to sample $\mathscr{M}$ 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

$$P_{\text{accept}} = \min\left(1, (L(\mathbf{m}_t)\rho_M(\mathbf{m}_t))/(L(\mathbf{m}_c)\rho_M(\mathbf{m}_c))\right). \quad (3)$$

- 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 requires 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* and 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

$$P_{\text{accept}} = \min(1, L(\mathbf{m}_t)/L(\mathbf{m}_c)). \quad (4)$$

- 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 the 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 an application of random function theory to characterize natural phenomena [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 [8, 14, 16, 25, 38, 43–45], and most are available through the SGeMS software package [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 two-point statistics and multiple-point statistics. Traditional geostatistical random function models rely on two-point statistics, where spatial variability is only described between pairs of two data locations, typically quantified by a

covariance model. The simplest two-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 two-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 two-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, Mustapha 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

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 $f_{\mathrm{M}}(m_1 \dots m_N)$ 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

$$f_{\mathrm{M}}(m_i | m_1 \dots m_{i-1}), \tag{5}$$

That the above procedure will actually generate a realization from $f_{\mathrm{M}}(m_1 \dots m_N)$ follows from the identity $f(s|t)\, f(t) = f(s,t)$, which (in the general multivariate case) yields

$$\begin{aligned} f_{\mathrm{M}}(m_{n+1} \dots m_N | m_1 \dots m_n) = {} & f_{\mathrm{M}}(m_{n+1} | m_1 \dots m_n) \\ & f_{\mathrm{M}}(m_{n+2} | m_1 \dots m_{n+1}) \\ & \vdots \\ & f_{\mathrm{M}}(m_N | m_1 \dots m_{N-1}). \end{aligned} \tag{6}$$

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 efforts have been made in the geostatistical community to efficiently compute conditional probability density functions, as in Eq. 5, based on two-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 $f_{\mathrm{M}}(m_1 \dots m_N)$. If we randomly select a model parameter, $m_i = m(\mathbf{u_i})$, compute the local conditional pdf

$$\rho_{\mathrm{M}}(m_i \,|\, m_1, m_2, \dots, m_{i-1}, m_{i+1}, \dots, m_N). \tag{7}$$

and draw a value from it, we get a new realization of the random field defined by $f_{\mathrm{M}}(m_1 \dots m_N)$. If this is repeated iteratively, it will be an application of the Gibbs sampler [13] and thus allow sampling from $f_{\mathrm{M}}(m_1 \dots m_N)$. Note that the Gibbs sampler only requires that a sample from ρ_{M} can be generated. The full conditional pdf of Eq. 7 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. 7, which can be done very effectively using methods developed for sequential simulation, Eqs. 5–6. 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 $f_{\mathrm{M}}(m_1 \dots m_N)$.

The sequential Gibbs sampler can be used as a prior sampler for the extended Metropolis algorithm [37]. 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 $\rho_{\mathrm{M}}(\mathbf{m}_{i\in U}\,|\,\mathbf{m}_{i\notin U})$. In the Appendix, it is demonstrated that this procedure, if applied iteratively, will sample the distribution $\rho_{\mathrm{M}}(\mathbf{m})$.

To draw a realization from $\rho_{\mathrm{M}}(\mathbf{m}_{i\in U}\,|\,\mathbf{m}_{i\notin U})$ we could compute $\rho_{\mathrm{M}}(\mathbf{m}_{i\in U}\,|\,\mathbf{m}_{i\notin U})$ 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 $\rho_{\mathrm{M}}(\mathbf{m}_{i\in U}\,|\,\mathbf{m}_{i\notin U})$ not the full conditional pdf itself. Instead, we can use the sequential simulation approach of Eq. 5, 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. 5.
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 three examples of using a random walk to generate realizations of a prior model defined by the spatial statistics of the training images in Fig. 1. We make use of the SNESIM algorithm to compute a realization of the conditional distribution Eq. 7 [45].

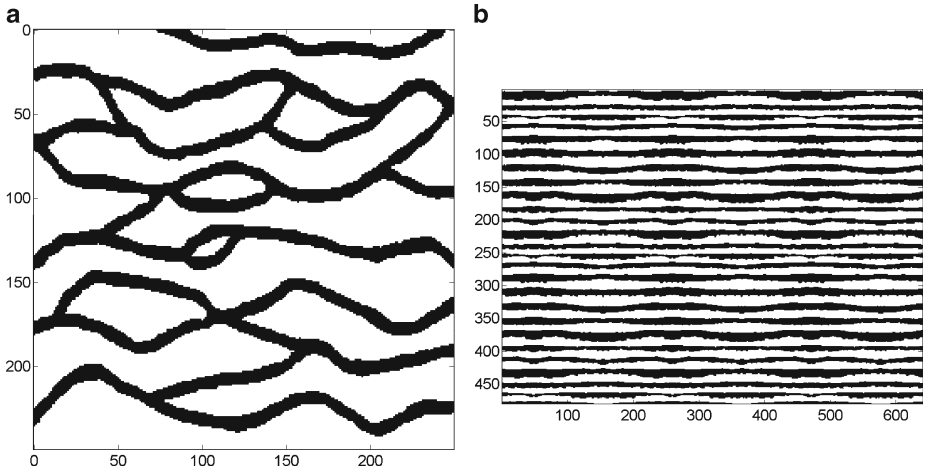


Fig. 1 Two different training images, reflecting a prior model based on higher-order spatial moments. **a** Channel-based training image from Strebelle (2002). **b** Pattern, used with permission from Ian Lynman

Figure 2a, b makes use of the channel-based training image of Fig. 1a, using a resimulation area, U, of 12×12 and 4×4 cells, respectively. Note that U does not need to take any specific shape. Any subset of model parameters can be considered. Figure 2c is based on the prior model associated with to the training image in Fig. 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 indicates the model parameter that is to be resimulated. In the following realization (to the right), the data within the gray area have been resimulated, and new area is selected for conditional resimulation. The final realization after six iterations of the sequential Gibbs sampler is the model to the right.

The training images of Fig. 2 are but two examples of a categorical two-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.

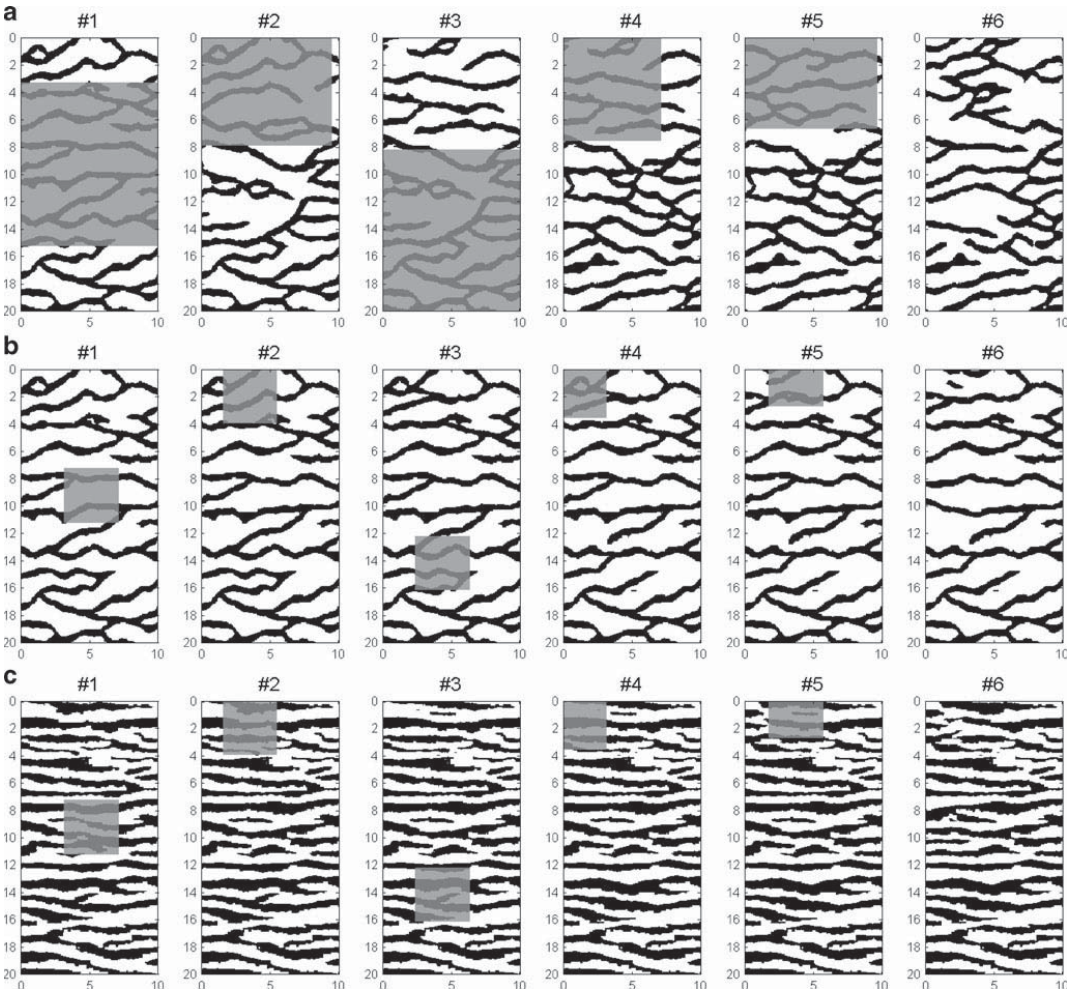


Fig. 2 Sequential Gibbs sampling using different areas of resimulation. **a** 12×12 cell resimulation using the training image in Fig. 1a. **b** 4×4 cell resimulation using the training image in Fig. 1a. **c** 4×4 cell resimulation using the training image in Fig. 1b. The *left column* is the starting model. The *right column* is the realization of the prior model after six iterations using the sequential Gibbs sampler

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.

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 describe 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 rest of the data in the block is simulated conditional to the edge data.

Without describing it in detail, Bosch [2] also seems to have made use of a technique similar to resimulation of one data point at a time as given in Eq. 5 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 41×41-pixel gray-scale image of Fig. 3 as a reference image. It has been generated using single normal equation simulation [45] using the training image in Fig. 1a. A Gaussian kernel with a horizontal range of 15 grid cells and a vertical range of six grid cells, see Fig. 3b, is used as a smoothing kernel to obtain a smoothed version of the reference gray-scale image in Fig. 3a, as seen in Fig. 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 (Fig. 3d). We now consider the 5×5 pixels indicated by circles in Fig. 3d as observed data.

Reconstructing the 41×41 gray-scale image (Fig. 3a) from the 5×5 pixel data can now be seen as a simple linear inverse problem, where the forward problem is described by

$$\mathbf{d} = \mathbf{Gm} \tag{8}$$

where $\mathbf{G}$ describes the smoothing kernel. A model of the noise can be described by a data covariance, $\mathbf{C}_{\mathrm{D}}$, 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, $\rho_{\mathrm{M}}^{\mathrm{TI}}$, $\rho_{\mathrm{M}}^{\mathrm{uncorr}}$, and $\rho_{\mathrm{M}}^{\mathrm{corr}}$ defined as

- $\rho_{\mathrm{M}}^{\mathrm{TI}}$
  The statistical model described by the reference training image (Fig. 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 $\rho_{\mathrm{M}}^{\mathrm{TI}}$.
- $\rho_{\mathrm{M}}^{\mathrm{uncorr}} \rightarrow N(\mathbf{m}_0, \mathbf{C}_{\mathbf{m}^{\mathrm{uncorr}}})$
  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, $\mathbf{C}_{\mathbf{m}^{\mathrm{uncorr}}} = \mathbf{I}\, 0.45^2$.
- $\rho_{\mathrm{M}}^{\mathrm{corr}} \rightarrow N(\mathbf{m}_0, \mathbf{C}_{\mathbf{m}^{\mathrm{corr}}})$
  From the reference image of Fig. 3a, we estimate an apparent covariance model, $\mathbf{C}_{\mathbf{m}^{\mathrm{corr}}}$, 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 $\rho_{\mathrm{M}}^{\mathrm{uncorr}}$, $\mathbf{m}_0$=0.28.

All three considered prior models are consistent with the spatial statistics of the reference image up to a certain order. $\rho_{\mathrm{M}}^{\mathrm{uncorr}}$ is the least informed prior model, reflecting only the lower-order mean and variance, while $\rho_{\mathrm{M}}^{\mathrm{corr}}$ also reflect the correct covariance. $\rho_{\mathrm{M}}^{\mathrm{TI}}$ is the

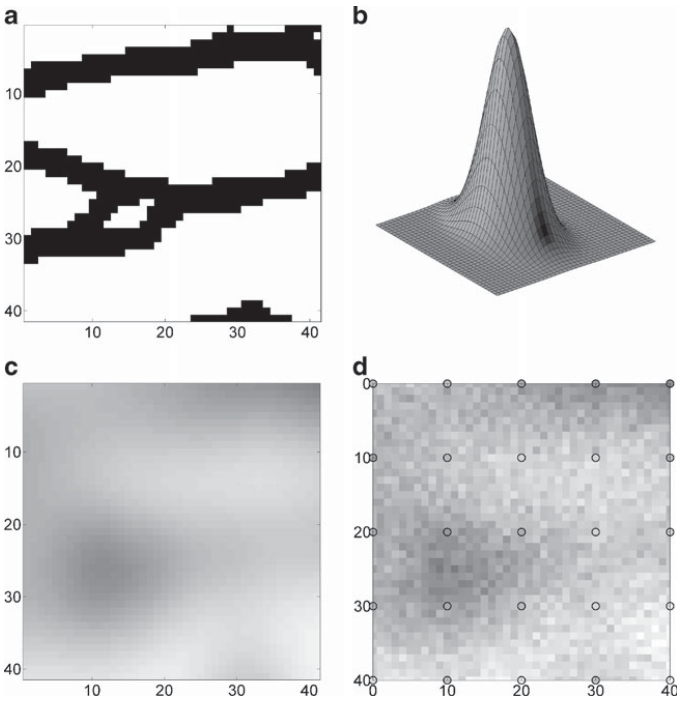


Fig. 3 **a** Reference image. **b** Gaussian smoothing kernel. **c** Smoothed reference image. **d** Smoothed reference image with noise. The 25-pixel values in the circles are used as data in the image reconstruction inversion

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. Assume that Gaussian-type prior models imply 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 characterized by the posterior mean and covariance, $N(\widetilde{\mathbf{m}}, \widetilde{\mathbf{C}}_{\mathrm{M}})$ (from [46]):

$$\widetilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_{\mathrm{M}}\mathbf{G}^t(\mathbf{G}\,\mathbf{C}_{\mathrm{M}}\,\mathbf{G}' + \mathbf{C}_{\mathrm{D}})^{-1}(\mathbf{d}_0 - \mathbf{G}\,\mathbf{m}_0) \quad (9)$$

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

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 $\rho_{\mathrm{M}}^{\mathrm{uncorr}}$ and $\rho_{\mathrm{M}}^{\mathrm{corr}}$, respectively, as seen in Fig. 4a, b.

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 two-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

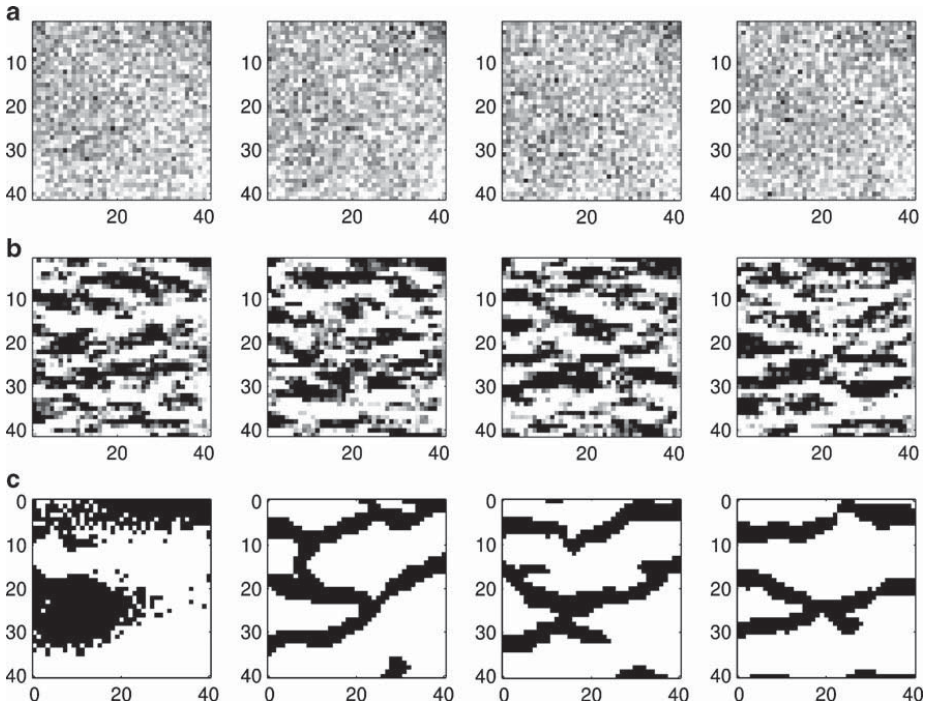


Fig. 4 Posterior samples using **a** ρ_M^{uncorr}, a Gaussian spatially uncorrelated prior model, **b** ρ_M^{corr}, a Gaussian spatially correlated prior model (an exponential covariance model with a horizontal range of 15 pixels and a vertical range of 6 pixels). **c** Initial model (leftmost) and three posterior realizations using the training image in Fig. 1a as prior model (ρ_M^{TI})

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 Fig. 3d where all pixel values below 0.28 is set to 0 and all pixel values above 0.28 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 Fig. 4c, first model).

The extended Metropolis algorithm is run for 10,000 iterations producing approximately 100 independent realizations of the posterior probability distribution. Three of these are shown in Fig. 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 Fig. 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 (Fig. 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, Fig. 5a, b shows this model with maximum posterior probability. Compared to the reference image of Fig. 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 (Fig. 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 Fig. 5c how the location of the channel is relatively sharply outlined, even though it is based on averaging

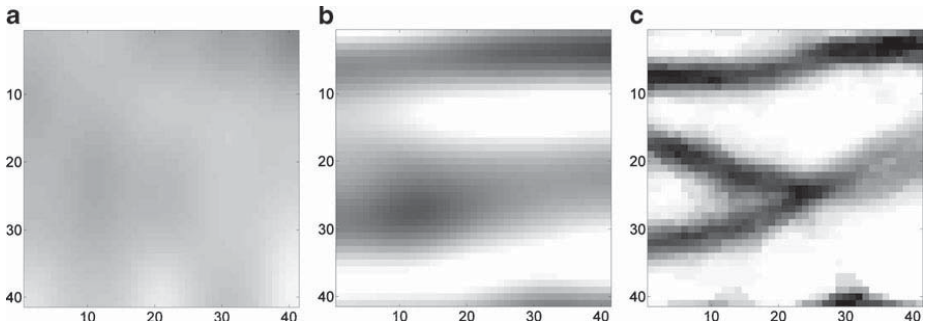


Fig. 5 Posterior statistics. **a** Posterior mean estimate using the uncorrelated Gaussian prior model, ρ_M^{uncorr}. **b** Posterior mean estimate using the correlated Gaussian prior model, ρ_M^{corr}. **c** Probability of a channel structure given the use of the training image-based prior model, ρ_M^{TI}

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 but it is also clear that adding consistent prior information for this case provides significantly sharper de-blurred images.

Typical application of the Metropolis algorithm for image reconstruction has been based on the classical Metropolis algorithms [1, 48]. This example shows 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. At 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 Fig. 1a, an unconditional realization (Fig. 6a) is generated using the SNESIM algorithm [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 Fig. 6a, and the corresponding arrival times are measured at the 40 receivers located to the right. This way 800 travel time data are recorded. Three percent Gaussian noise is added to the synthetic travel time data (Fig. 6b) and used as observed data, $\mathbf{t}_{obs}$. To compute the observed travel times, we use a

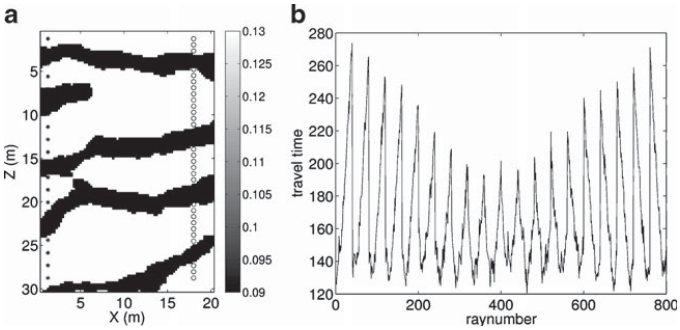


Fig. 6 **a** Reference velocity model and location of sources (*asterisk*) and receivers (*circle*). *Black* channel structures have a velocity of 0.09 m/ns. The background velocity (*white*) has a velocity of 0.13 m/ns. **b** Calculated first arrival travel time for waves traveling between sources and receivers. Three percent normally distributed noise was added to the travel times

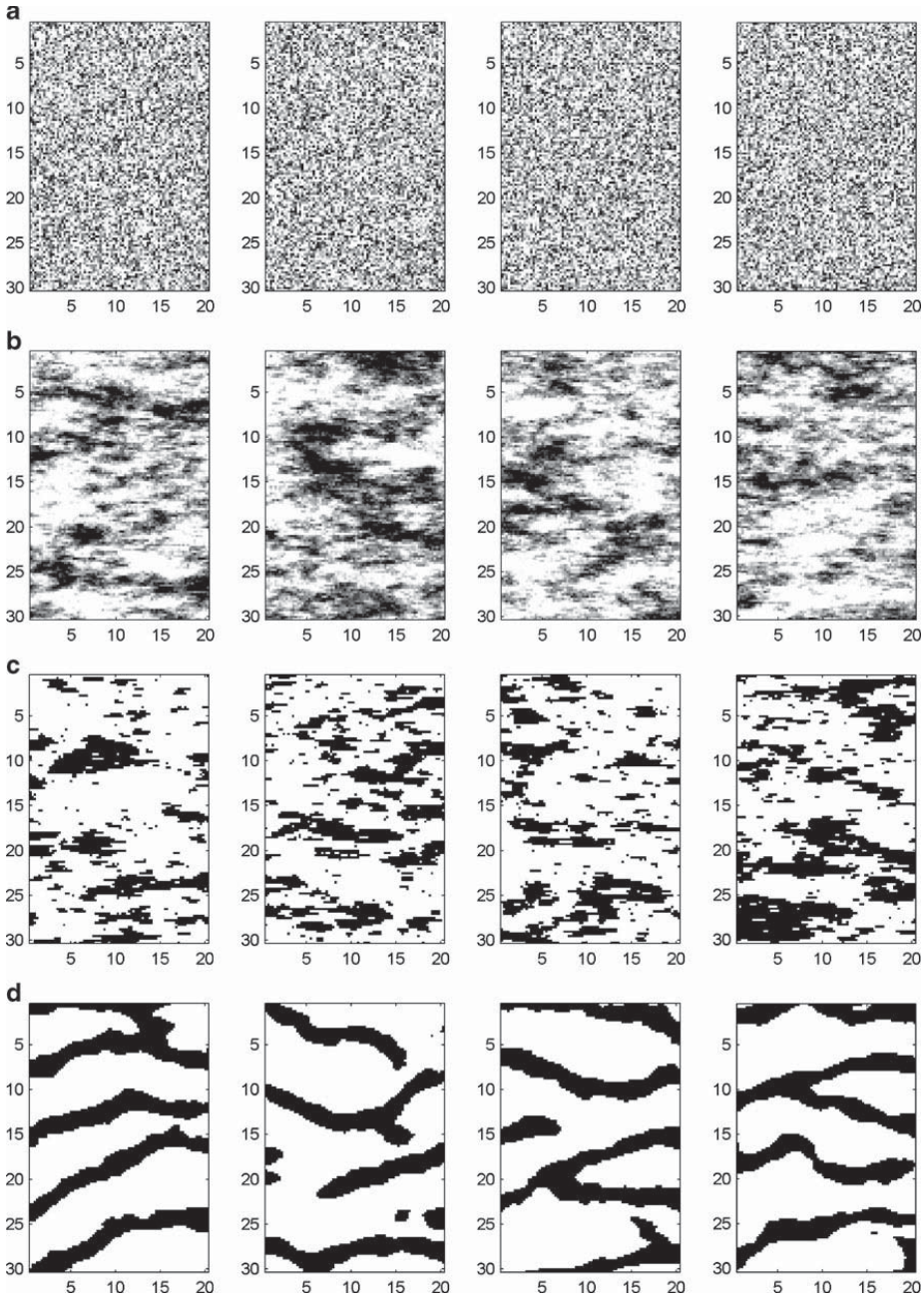


Fig. 7 Four realizations of the prior model of type **a** ρ_M^{nugget}, **b** ρ_M^{sgsim}, **c** ρ_M^{dssim}, and **d** ρ_M^{TI} prior models

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, $\mathbf{t}_{\text{est}}(\mathbf{m})$, is computed as

$$L(\mathbf{m}) = \exp(-\frac{1}{2}(\mathbf{t}_{\text{est}}(\mathbf{m}) - \mathbf{t}_{\text{obs}})' \mathbf{C}_{\text{D}}^{-1} (\mathbf{t}_{\text{est}}(\mathbf{m}) - \mathbf{t}_{\text{obs}})) \tag{11}$$

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

We consider seven different prior models based on both two-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 Fig. 1a. Thus, all considered prior models are consistent with this lower-order statistics of the training image.

The first five prior models are based on two-point random models. A pure nugget model assumes no spatial correlation, and thus, all model parameters are a priori considered uncorrelated, $\rho_{\text{M}}^{\text{nugget}}$. Two models, $\rho_{\text{M}}^{\text{Gau}_1}$ and $\rho_{\text{M}}^{\text{Gau}_8}$, are based on a Gaussian covariance

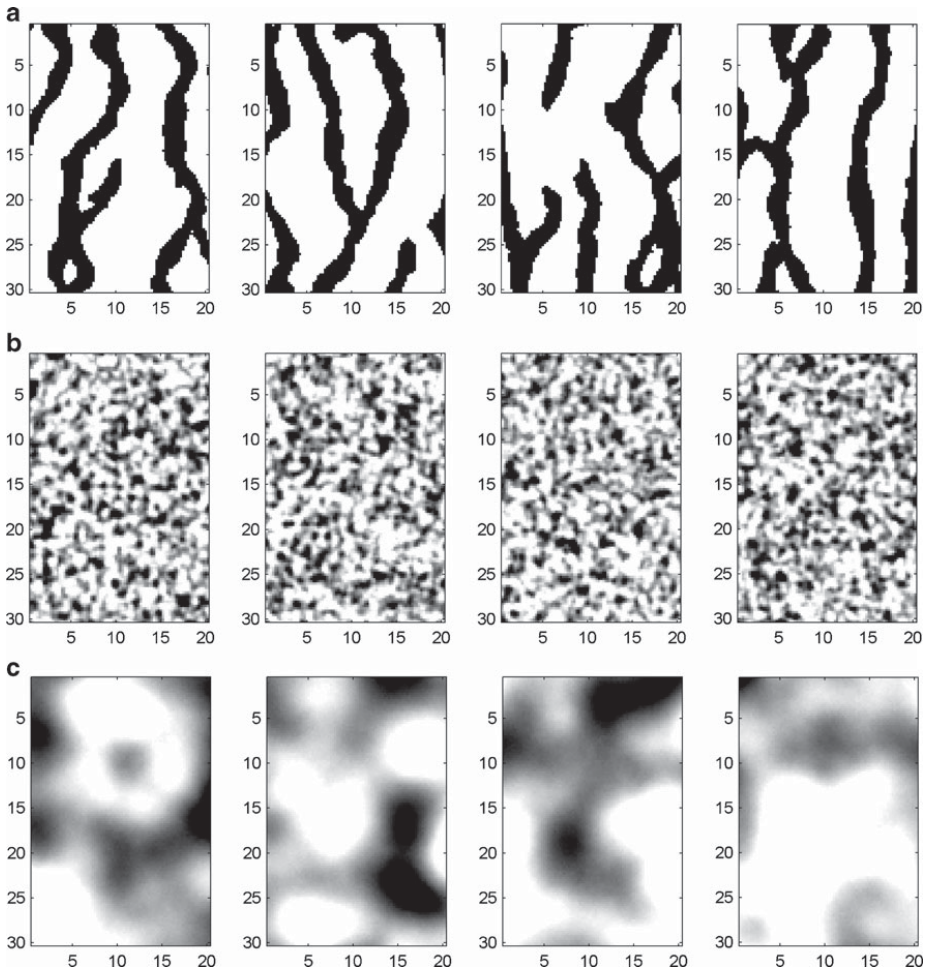


Fig. 8 Four realizations of the prior model of type **a** $\rho_{\text{M}}^{\text{TI}_{90}}$, **b** $\rho_{\text{M}}^{\text{Gau}_1}$, and **c** $\rho_{\text{M}}^{\text{Gau}_8}$

model with an isotropic range of 1 and 8 m, respectively. The ρ_M^{sgsim} prior is based on a covariance model inferred from the training image in Fig. 1a, as an exponential covariance model with a horizontal range of 6.6 m and a vertical range of 2.2 m. 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 Fig. 1a. The last prior, $\rho_M^{TI_{90}}$, is based on the multiple-point statistical model inferred from the training image rotated 90° clockwise. The four 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 and 8 show four unconditional realizations, i.e., a sample, of the seven considered prior models. For each of these prior models, an extended Metropolis algorithm is run for 35,000 iterations. The VISIM [18] and SNESIM [45] programs have been used to perform sequential simulation, needed for applying the sequential Gibbs sampler, for two-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. Figures 10 and 11 show the current model at iteration 20,000, 25,000, 30,000, and 35,000.

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 Fig. 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., [4, 7, 12].

It is clear from Fig. 9 that the burn-in phase has not been completed in the considered 35,000 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 20,000 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 ∞, 4,000, 3,800, and 1,000, 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

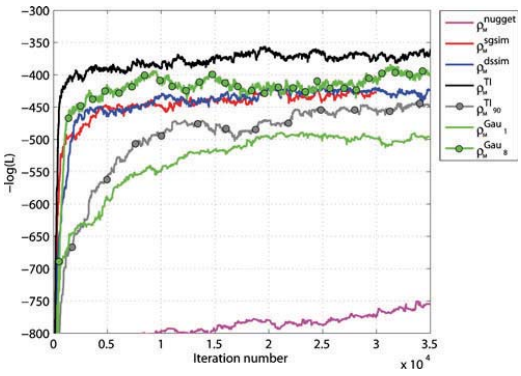


Fig. 9 Negative log-likelihood as a function of iteration number for different choices of prior model

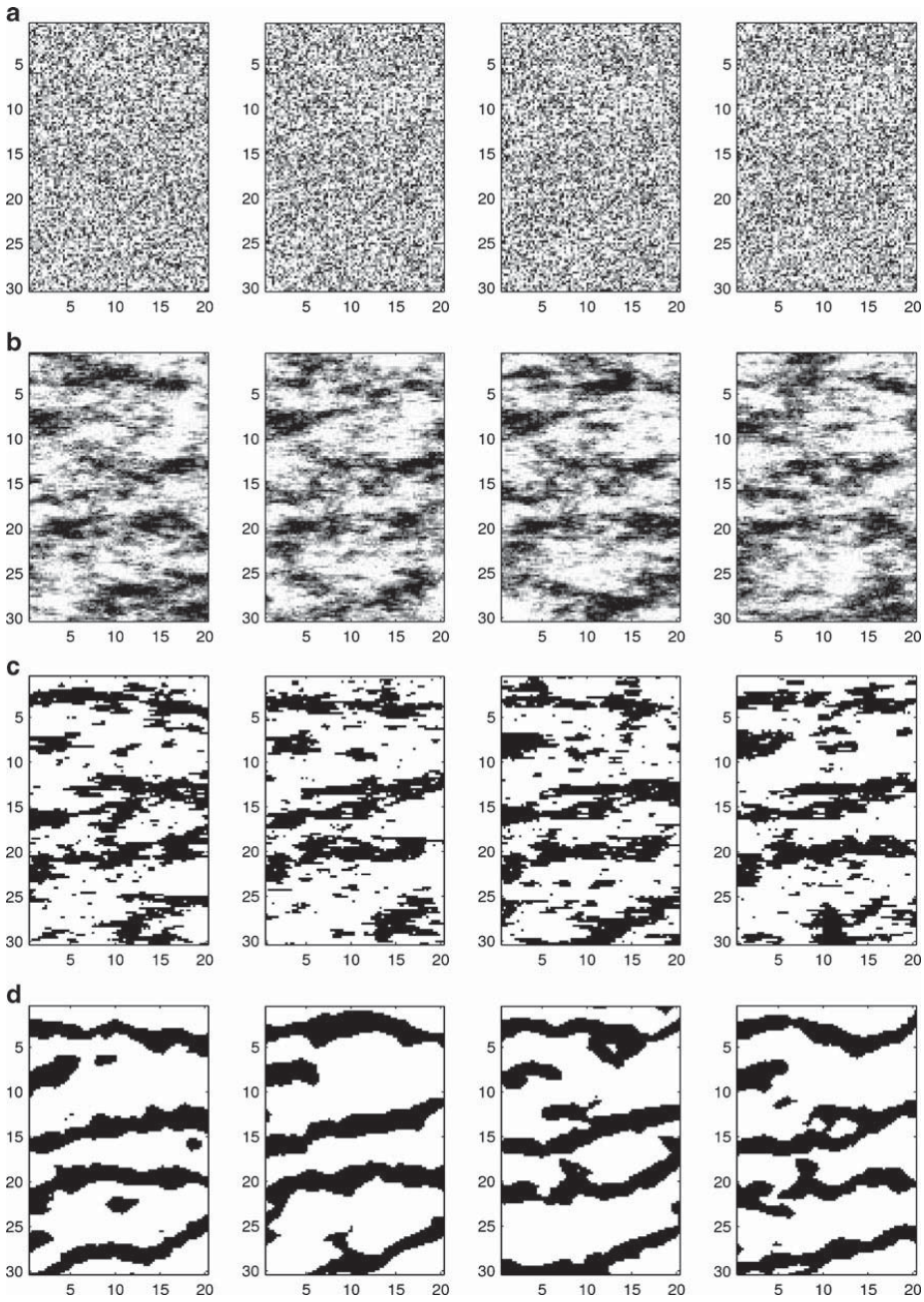


Fig. 10 Current model at iteration number 20,000, 25,000, 30,000, and 35,000 using the **a** ρ_M^{nugget}, **b** ρ_M^{sgsim}, **c** ρ_M^{dssim}, and **d** ρ_M^{TI} prior models, respectively

of the sampling efficiency of the extended Metropolis algorithm reveals that, on average, 15,000, 6,000, and 2,500 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 six 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 illustrate 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 four independent realizations of ρ_M^{TI}. It is thus a visualization of the prior assumptions inherent in ρ_M^{TI}. Figure 10d shows a corresponding sample in the form of four independent realizations of the posterior pdf, σ_{TI}, and is thus a graphical representation of the state of information of

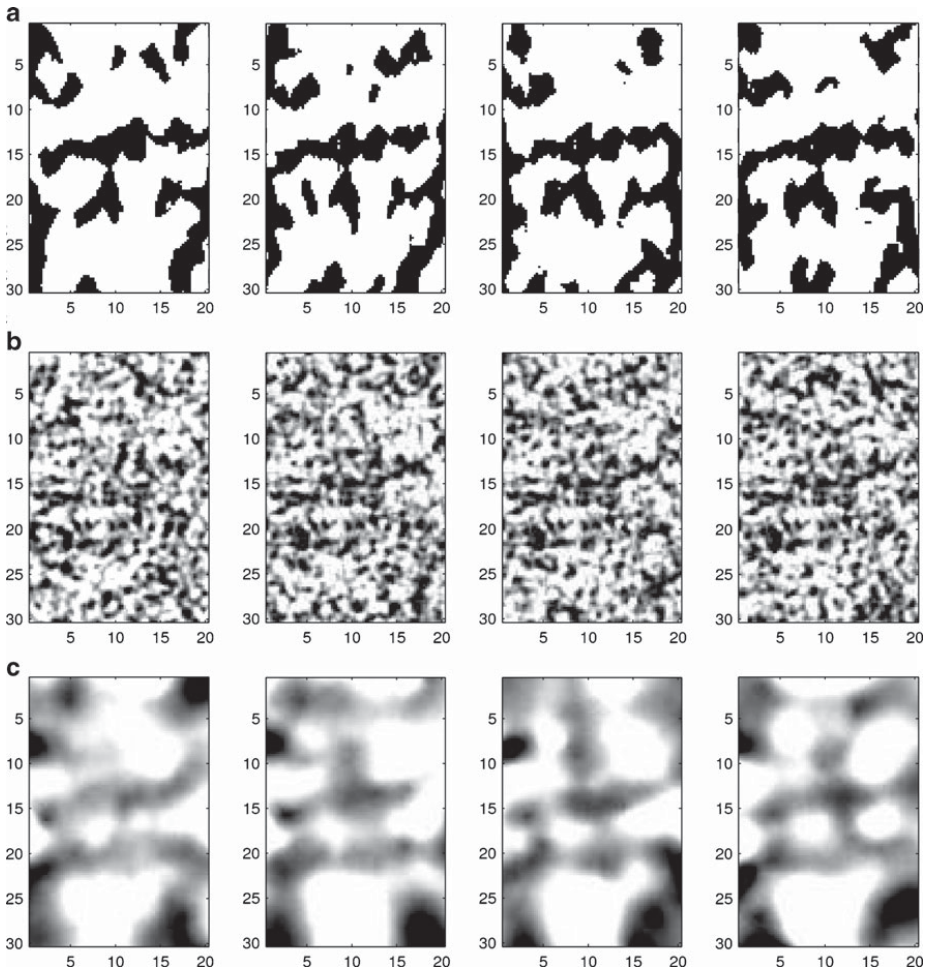


Fig. 11 Current model at iteration number 20,000, 25,000, 30,000, and 35,000 using the **a** $\rho_M^{TI_{90}}$, **b** $\rho_M^{Gau_1}$, and **c** $\rho_M^{Gau_8}$ prior models

the model parameters obtained by merging the states of information as given by the prior and the likelihood, Eq. 2. The differences between Figs. 7d and 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 Figs. 10 and 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 noninformative 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 maximum posterior model. 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 model with maximum posterior probability. 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 impossible 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 two-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.

Acknowledgments We thank DONG E&P, Denmark, for financial support. Klaus Mosegaard was supported by The Danish Council for Independent Research—Natural Sciences (FNU), Grant 272-08-0568. We thank Ian Lynam (http://www.ianlynam.com/) for permission to use his pattern (Fig. 1b). SNESIM and SGeMS were used for multiple-point-based simulation, while VISIM and mGstat (http://mgstat.sourceforge.net) were used for second-order statistical simulation. We thank two anonymous reviewers for their thorough evaluation and constructive suggestions that helped improve this manuscript.

Appendix: The Gibbs sampler with multiple parameter perturbations

Consider a Gibbs sampler with multiple parameter perturbations, perturbing in iteration k all parameters belonging to resimulation 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:

$$P_k(\mathbf{m}_i|\mathbf{m}_j) = \begin{cases} \frac{p(\mathbf{m}_i)}{\sum_{\mathbf{m}_k \in \mathcal{N}_j^k} p(\mathbf{m}_k)} & \mathbf{m}_i \in \mathcal{N}_j^k \\ 0 & \text{otherwise} \end{cases} \tag{A1}$$

where $\mathcal{N}_j^k$ is the set of all points that are identical to $\mathbf{m}_j$ in parameters not belonging to the resimulation 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 $P_k(\mathbf{m}_i|\mathbf{m}_j)$:

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

where we have used that $\mathcal{N}_j^k = \mathcal{N}_i^k$. This property, called *detailed balance*, 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 toward $p(\mathbf{x})$ regardless of the starting distribution. The two conditions are:

1. *Aperiodicity*. 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. *Irreducibility*. It is possible to go from any point $\mathbf{m}_j$ to any other point $\mathbf{m}_i$ in $\mathscr{M}$, 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.

References

1. Besag, J.: On the statistical analysis of dirty pictures. J. R. Stat. Soc., B **48**(3), 259–302 (1986)
2. Bosch, M.: Lithologic tomography: from plural geophysical data to lithology estimation. J. Geophys. Res. **104**(B1), 749–766 (1999)
3. Box, G., Tiao, G.: Bayesian Inference in Statistical Analysis. Wiley, New York (1973)
4. Brooks, S., Gelman, A.: General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. **7**, 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. Math. Geol. **38**(1), 81–100 (2006)
7. Cowles, M., Carlin, B.: Markov chain Monte Carlo convergence diagnostics: a comparative review. J. Am. Stat. Assoc. **91**, 883–904 (1996)
8. Deutsch, C.V., Tran, T.T., Xie, Y.: An Approach to Ensure Histogram Reproduction 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 Probl. Sci. Eng. **16**(7), 865–884 (2008)
10. Fu, J., Gómez-Hernández, J.: A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling. Math. Geosci. **41**(2), 105–128 (2009)
11. Gelman, A., Roberts, G., Gilks, W.: Efficient metropolis jumping rules. In: Bernardo, J., Berger, K., Dawid, A., Smith, A. (eds.) Bayesian Statistics, vol. 5, pp. 599–608. Clarendon, Oxford. (1996)
12. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**, 457–472 (1992)
13. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
14. Gomez-Hernandez, J., Journel, A.: Joint sequential simulation of multi-Gaussian fields. In: Soares, A. (ed.) Geostatistics Troia, vol. 92, pp. 85–94. Kluwer Academic, Dordrecht (1993)
15. Goovaerts, P.: Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series. Oxford University Press, Oxford (1997)
16. Guardiano, F., Srivastava, R.: Multivariate geostatistics: beyond bivariate moments. In: Soares, A. (ed.) Geostatistics-Troia, vol. 1, pp. 133–144. Kluwer Academic, Dordrecht (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. Comput Geosci **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: Ortiz, J.M., Emery, X. (eds.) VIII International Geostatistics Congress, vol. 1, pp. 329–338. Mining Engineering Department, University of Chile, Santiago (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: Bernardo, J., DeGroot, M., Lindley, D., Smith, A. (eds.) Bayesian Statistics, vol. 2, pp. 329–360. Elsevier, Amsterdam (1985)
23. Jaynes, E., Bretthorst, G.: Probability Theory: the Logic of Science. Cambridge University Press, Cambridge (2003)
24. Journel, A., Zhang, T.: The necessity of a multiple-point prior model. Math. Geol. **38**(5), 591–610 (2006)
25. Journel, A.G.: Modeling uncertainty: some conceptual thoughts. In: Dimitrakopoulos, R. (ed.) Geostatistics for the Next Century, pp. 30–43. Kluwer, Dordrecht (1994)
26. Journel, A.G., Huijbregts, C.J.: Mining Geostatistics. Academic, Amsterdam (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 simulated annealing. Science **220**(4598), 671–680 (1983)
29. Le Ravalec-Dupin, M., Noetinger, B.: Optimization with the gradual deformation method. Math. Geol. **34**(2), 125–142 (2002)
30. Mariethoz, G., Renard, P., Caers, J.: Bayesian inverse problem and optimization with iterative spatial resampling. Water Resour. Res. **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 Probl. **14**, 405 (1998)
33. Mosegaard, K.: Monte Carlo Analysis of Inverse Problems. University of Copenhagen, Copenhagen (2006)
34. Mosegaard, K.: Quest for consistency, symmetry and simplicity—the legacy of Albert Tarantola. Geophysics **76**, W51–W61 (2011)
35. Mosegaard, K., Sambridge, M.: Monte carlo analysis of inverse problems. Inverse Probl. **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. J. Geophys. Res. **102**(B2), 2969–2981 (1997)
37. Mosegaard, K., Tarantola, A.: Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. **100**(B7), 12431–12447 (1995)
38. Mustapha, H., Dimitrakopoulos, R.: HOSIM: a high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns. Comput Geosci (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: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, Cambridge (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, Philadelphia (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. Scand. J. Statist. **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. Comput Geosci **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. J. Geophys. Res. **103**(B4), 7187–7210 (1998)
52. Zhang, T.: Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics. Earth Sci Front **15**(1), 26–35 (2008). doi:10.1016/S1872-5791(08)60016-0, 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, pp 24–29 (2009). doi:ieeecomputersociety.org/10.1109/ICACC.2009.33

Appendix A3:

Monte Carlo full waveform inversion of tomographic crosshole data using complex geostatistical a priori information

Authors:

Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard

Published in:

Expanded abstract of Society of Exploration Geophysicists' 80th Annual Meeting (SEG2010)

Denver, Colorado, USA

18 – 22 October 2010

Monte Carlo full waveform inversion of tomographic crosshole data using complex geostatistical a priori information

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

Summary

This paper presents a Monte Carlo full waveform inversion strategy based on a Bayesian formulation of the inverse problem. Existing full waveform inversion strategies often relies on a migration based approach, which suffers from lack of uncertainty estimates. Using a Bayesian approach, the solution to the inverse problem is formulated as an a posteriori probability density. We demonstrate that samples from the solution to the full waveform inverse problem can be obtained using the extended Metropolis algorithm in conjunction with complex a priori information. The a priori information is described by a training image using a geostatistical algorithm. A posteriori samples from the solution to the inverse problem provide a means of obtaining resolution analysis of the solution. The suggested strategy is tested on synthetic crosshole full waveform ground penetrating radar (GPR) data, but is equally well applicable to seismic waveform data. The forward problem is solved using finite-difference time-domain calculations of Maxwell's equations. To our knowledge this is the first example of performing full waveform inversion using the extend Metropolis Algorithm and, in this way, provide an uncertainty estimate of a tomographic full waveform inverse problem.

Introduction

Ground penetrating radar (GPR) crosshole tomography is a popular method used to obtain tomographic images of near-surface geological structures and geophysical parameters. The crosshole GPR experiment involves a transmitting radar antenna (20 MHz – 1 GHz, (see Reynolds, 1997)) lowered into a borehole and a receiving antenna placed in an adjacent borehole. The boreholes are typically separated by a distance of 5m – 20m and are 10 m – 20 m deep. The transmitting antenna is kept fixed in one borehole while the signal is recorded at multiple locations in the opposite borehole. This procedure is repeated for multiple transmitter positions until the inter borehole section has been covered with measurements. For a corresponding seismic experiment the source pulse frequency is typically in the order of 500Hz – 1000 Hz (e.g. Paasche et al., 2006; Belina et al., 2009).

Ray-based inversion of first arrival travel times and amplitudes of GPR or seismic data provide estimates of the signal velocity and attenuation distribution of the inter-borehole region. Unfortunately, ray-theory is based on a high-frequency approximation and accounts only for a small part of the information content of the full waveform signal. Therefore, tomographic images obtained using ray-theory are limited in resolution.

Ernst et al. (2007a,b) introduced and applied an algorithm for inversion of tomographic crosshole GPR data. This code was modified by Belina et al. (2009) and applied to seismic waveform data. Ernst et al. (2007a,b) and Belina et al. (2009) demonstrated that sub-wavelength features can be resolved by properly modelling the wave propagation and, hence, utilizing the full information content of the waveform data. The algorithm introduced by Ernst et al. (2007a,b) relies on the migration-based full waveform inversion concept introduced by Tarantola (1984). This method is desirable because it, in relatively few iterations, is able to infer geophysical parameters in order to minimise a misfit function between the observed and modelled waveforms. However, the migration based approach has some limitations: Firstly, the method depends on a good initial guess of the large scale features of the solution in order to ensure convergence. Secondly, when using this method in a tomographic crosshole setup, a dense receiver and transmitter coverage is needed in order to avoid numerical artefacts along the boreholes. Finally, the method is based on an optimization algorithm and does not provide any uncertainty estimate of the solution.

Here, we formulate the GPR crosshole problem as a probabilistic, Bayesian, full waveform inverse problem. We use the extended Metropolis Algorithm (Mosegaard and Tarantola, 1995) to obtain samples of the a posteriori pdf. In this way, not only a single estimate, but multiple realizations of tomographic images which all honour the data uncertainties and a priori information are provided. This, in turn, provides a means of performing resolution analysis of the solution (Mosegaard, 1998). The a priori information about the problem is controlled through a geostatistical algorithm. This approach allow us to incorporate complex a priori information described by both 2-point and multiple-point statistics. In this study we limit the example only to concern a priori information described by a training image (i.e. multiple-point statistics). We demonstrate how geostatistical formulated a priori information serves as a guide in the initial part of the inversion procedure. In this way convergence of the suggested full-waveform inversion algorithm becomes independent of the initial model.

Monte Carlo full waveform inversion

Methodology

Consider that the subsurface can be represented by a discrete set of model parameters, $\mathbf{m}$, and that a data set, $\mathbf{d}$, of indirect observations of the model parameters is provided. The model parameters describe some physical properties of the subsurface that influences the data observations. Hence, the forward relation between the model parameters (i.e. the model) and the data observations can be expressed as (e.g. Tarantola, 2005):

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

where g is a linear or non-linear mapping operator which often relies on a physical law. Here, the forward relation in equation (1) is given as a finite-difference time-domain solution of Maxwell's equations. However, any numerical wave propagation modelling strategy for GPR or seismic signals can be applied. The inverse problem is to infer information about the model parameters based on a set of observations, a priori information about the model, and the forward relation between the model and the data observations.

In a Bayesian formulation the solution to the inverse problem is given as an a posteriori probability density, which can be formulated as (e.g. Tarantola, 2005):

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

where k is a normalization constant, $\rho_M(\mathbf{m})$ is the a priori probability density, 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 modelled data explains the observed data given a data uncertainty. Hence, the a posteriori probability density describes the probability that a certain model is a solution to the inverse problem.

A highly nonlinear inverse problem refers to the case where the a priori probability density is far from being Gaussian or the forward relation between the model and data are far from being linear. In the case of full waveform inversion the forward relation is expected to be highly nonlinear. Moreover, the a priori information described by a training image is highly non-Gaussian.

The extended Metropolis algorithm is a versatile tool which, in particular, is useful to obtain samples from solutions to non-linear inverse problems using arbitrarily complex a priori information. The minimum requirement of the algorithm is; 1) a "black box" algorithm that is able to sample the a priori probability density and, 2) a "black box" algorithm that is able to compute the likelihood for a given set of model parameters. The flowchart of the extended Metropolis algorithm is as follows: 1) The a priori sampler proposes a sample, $\mathbf{m}_{propose}$, from the a priori probability density, which is a perturbation of a previous accepted model, $\mathbf{m}_{accept}$. 2) The proposed sample is accepted with the probability (known as the Metropolis rule):

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

3) If the proposed model is accepted, $\mathbf{m}_{propose}$ is a sample of the a posteriori probability and $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{accept}$. Otherwise $\mathbf{m}_{propose}$ is rejected. 4) The procedure is continued until a desirable number of models have been accepted.

In this study the algorithm that provides the a priori information is the Single Normal Equation SIMulation (snesim) algorithm, which is a fast geostatistical algorithm that produces samples (conditional or unconditional) from an a priori probability density defined by a training image for a relatively low number of categorical values (Strebelle, 2002). Hansen et al. (2008) suggest a strategy termed perturbed simulation, which is capable of producing perturbations of spatial distributions using geostatistical algorithms. Thus, perturbed simulation serves as a "black box" that produces samples of a priori probability densities described by both two-point and multiple-point statistics. The flow of this algorithm is as follows: 1) An initial unconditional sample of the a priori probability density (here defined by a training image) is provided. 2) A subarea of the sample is randomly chosen. 3) The model parameters within this area are set to unknown. 4) The unknown model parameters are resimulated conditional to the rest of the model parameters using a geostatistical algorithm (here snesim) and a perturbation is obtained. 5) This procedure is repeated in order to obtain multiple samples of the a priori probability density.

The size of the perturbation area governs the exploratory nature of the Metropolis algorithm. The size of the perturbation area is chosen subjectively. In the extreme case where the area covers the entire model the outcome of the perturbed simulation algorithm is uncorrelated to the previous model. Contrary, if the area only constitutes a single model parameter, the perturbed model is highly correlated with the initial model. According to the metropolis rule a small perturbation area results in a proposed model that is more probable of being accepted as compared to a proposed model obtained using a larger perturbation area. Therefore, the perturbation area should

be chosen carefully in order to ensure an efficient algorithm. Gelman et al. (1996) found that the acceptance rate should be around 23% for high-dimensional distributions. For large acceptance rates the algorithm is exploring the a posteriori probability density too slowly. On the other hand, for smaller acceptance rates too many computationally expensive trials are performed. Therefore, we suggest to automatically change the size of the perturbation area while running the algorithm such that a certain acceptance rate is maintained. A constant acceptance rate results in a larger perturbation area in the burn-in period than in the subsequent sampling period. This effect is beneficial because the algorithm needs to perform large perturbations in the initial part in order to find models of large probability and, hence, produce representative samples of the a posteriori probability density.

Finally, the likelihood function is defined as a Gaussian distribution:

$$L(\mathbf{m}) = k \exp\left(-\frac{1}{2}\sum_{i=1}^{N}(g(\mathbf{m})^i - d_{obs}^i)/\sigma^2\right), \qquad (4)$$

where $g(\mathbf{m})^i$ represents the amplitude of the individual sample points of all the simulated waveforms obtained through equation (1) (i.e. the FDTD algorithm) and d_{obs}^i are the sample points of the observed waveform data. σ is the standard deviation of the expected amplitude uncertainty of the waveform data.

Results and discussion

Figure 1 shows a training image that mimic a matrix of clay with embedded channels of unconsolidated sand. Electromagnetic signals in near surface sediments are sensitive to the dielectric permittivity and the electrical conductivity of the materials. In this study we limit ourselves only to consider the influence of the dielectric permittivity, which is primarily governing the phase velocity of the signal. Water saturation of clay is often high compared to sandy deposits. Therefore, the dielectric permittivity of the clay is set to a relative dielectric permittivity of $\varepsilon_r \approx 4{,}57$ (0,14m/ns) and the permittivity of the sand channels is set to $\varepsilon_r \approx 2{,}75$ (0,18m/ns) (e.g. Topp et al., 1980). Figure 2 (left) is the synthetic reference to be considered and is, at the same time, an unconditional sample of the training image obtained using snesim. The electrical conductivity is set to a constant value of 3 mS/m and is, in the following, assumed known.

A full waveform synthetic data set is calculated using the FDTD algorithm. A Ricker wavelet with a central frequency of 100 MHz is used as source pulse. The source pulse is assumed known during the inversion. The transmitter and receiver positions are separated by 2 m and 0.25 m, respectively (see figure 2 left). Data acquired with a transmitter-receiver angle larger than 45 degrees from horizontal are omitted since, in practice, these data are violated by effects of wave guiding in the boreholes (cf. Peterson, 2001). This leads to a total of 248 data observations (i.e. recorded waveforms).

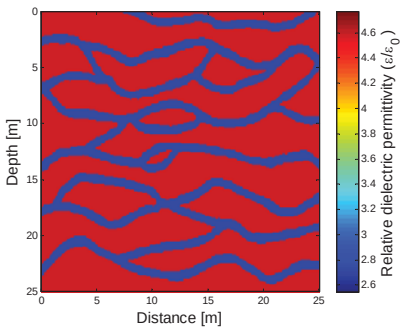


Figure 1. Training image which mimic sandy channel structures embedded in a matrix of clay deposits.

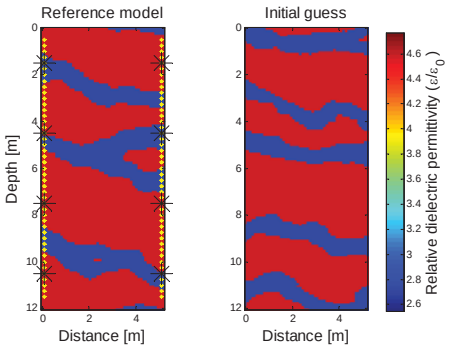


Figure 2. Left) Synthetic reference model. Black asterisks show transmitter positions and the yellow dots show receiver positions. Right) The initial model used as input for the inversion.

Noise is subsequently added to the data by performing a random phase shifting of the synthetic waveforms. The phase shift is normal distributed with zero mean and a standard deviation of 0.4 ns since this is a typical magnitude found in GPR travel time data (e.g. Looms et al, in press). The phase shift results in an amplitude

uncertainty with a standard deviation of 10^{-3}. Accordingly, the standard deviation of the data uncertainties, σ, is set to 10^{-3}. The standard deviation of the amplitude noise is indicated by the red error bar in figure 3 and compared with two waveforms recorded at 0 degrees (short offset / high amplitude) and 45 degrees transmitter-receiver angles (long offset / low amplitude), respectively.

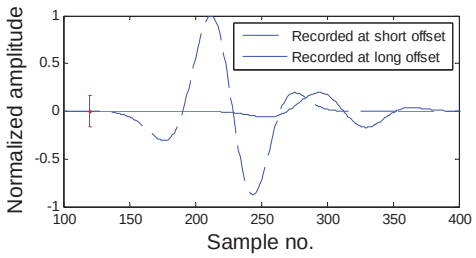


Figure 3. Dashed line is a waveform recorded at short transmitter-receiver offset (0 degrees). Solid line is a waveform recorded at long offset (45 degrees). The height of the red errorbar indicates 2 times the standard deviation of the noise added to the data.

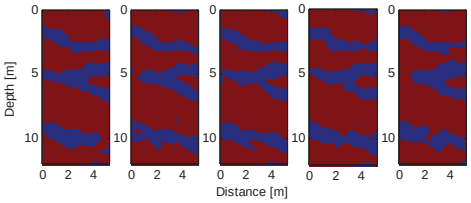


Figure 4. Five statistically independent samples from the a posteriori probability density of the full waveform inverse problem.

The initial model used for the Metropolis algorithm is chosen as an unconditional sample of the training image using a different random seed than for the reference model (see figure 2 right). Burn-in was reached after 2000 accepted models. Hereafter samples accepted by the Metropolis rule are representative samples of the a posteriori probability density. Figure 4 shows the 2000th, 6000th, 1000th, 14000th, and 18000th accepted sample using a priori information defined by the training image (figure 1) and Gaussian data uncertainty (equation 4). Only a slight deviation between the individual samples is seen, which indicates little a posteriori model uncertainty. Figure 5 shows the a posteriori mean and variance based on 18000 samples from the a posteriori probability density after the burn-in period. From the a posteriori variance it is seen that the overall structures of the model is recovered (variance close to or equal to zero) whereas the higher variances along the edges between the clay and sand deposits are subjected to uncertainty. A comparison between the reference model (figure 2 left) and the mean of the samples (figure 4 right) confirms that the waveform inversion is able to recover the sand structures very well. Moreover, it should be noted that the algorithm is able to reach these results even though it is initiated in a model uncorrelated with the reference model (compare figure 2 right and left).

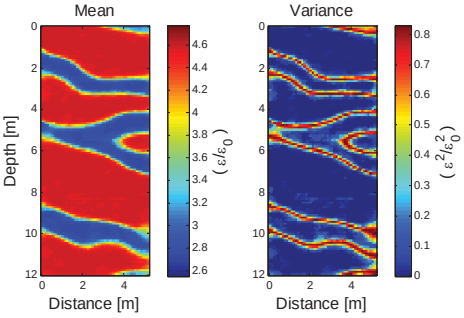


Figure 5. Mean (left) and variance (right) of 18000 samples drawn from the a posteriori probability density of the solution to the full waveform inverse problem.

The suggested Monte Carlo inversion strategy is preferable in that it allows for arbitrary antennae geometry. Furthermore, complex a priori inversion can be included using the perturbed simulation algorithm. Finally, a full data covariance matrix can be included in order to account for correlated data errors, which are often present in tomographic inverse problems (e.g. Maurer and Musil, 2004; Cordua et al., 2008, 2009). However, note that this exhaustive sampling strategy needs substantially more computationally expensive forward calculations compared to the traditional migration based approach.

Conclusions

We have demonstrated the potential of producing samples of the solution to a tomographic full waveform inverse problem using the extended Metropolis algorithm with complex a priori information. The methodology provides a means of evaluating the a posteriori uncertainty, which is not provided using optimisation based strategies for full waveform inversion. Moreover, the present approach is robust with regard to the initial guess of the solution and the transmitter-receiver density. Finally, the extended Metropolis Algorithm is flexible regarding the choice of a priori information and specification of data uncertainties.

References

Belina, F. A., J. R. Ernst, and K. Holliger, 2009, Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy *c*-means cluster analyses: Journal of Applied Geophysics, **68**, 85 – 94.

Cordua, K. S., M.C. Looms, and L. Nielsen, 2008, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data: Vadose Zone Journal, **7**, 263 – 271.

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.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007a, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, J53 – J64.

Ernst, J.R., H. Maurer, A. G. Green, and K. Holliger, 2007b, Full-waveform inversion of crosshole radar data based on 2- D finite-difference time-domain solutions of maxwell's equations: IEEE Transactions on Geoscience and Remote Sensing, **45**, 2807 – 2828.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient Metropolis jumping rules: Bayesian statistics, **5**, 599 – 607.

Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2008, Using geostatistics to describe complex a priori information for inverse problems, in J. M. Ortiz and X. Emery, eds., Geostatistics 2008 Chile, vol. 1.

Maurer, H.R. and M. Musil, 2004, Effects and removal of systematic errors in crosshole georadar attenuation tomography: Journal of Applied Geophysics, **55**, 261 -260.

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: Inverse Problems, **14**, 405 – 426.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of geophysical research, **100**, 431 – 447.

Paasche, H., J. Tronicke, K. Holliger, A. G. Green, and H. Maurer, 2006, Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses: Geophysics, **71**, H33–H44.

Peterson, J.E., 2001, Pre-inversion correction and analysis of radar tomographic data: Journal of Environmental and Engineering Geophysics, **6**, 1 – 18.

Reynolds, J.M., 1997, An introduction to applied and environmental geophysics: John Wiley and Sons Ltd. Pp.

Strebelle, S., 2002, Conditional simulation of somplex geological structures using multiple-point statistics: Matehmatical geology, **34**, 1 – 21.

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

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259 – 1266.

Topp, G.C., J. L. Davis, and A. P. Annan, 1980, Electromagnetic determination of soil water content: Measurements coaxial transmission lines: Water Resources Research, **16**, 574–582.

Appendix A4:

Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information

Authors:

Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard

Published in:

Geophysics

GEOPHYSICS, VOL. 77, NO. 2 (MARCH-APRIL 2012); P. H19–H31, 11 FIGS.
10.1190/GEO2011-0170.1

Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information

Knud Skou Cordua[1], Thomas Mejer Hansen[1], and Klaus Mosegaard[1]

ABSTRACT

We present a general Monte Carlo full-waveform inversion strategy that integrates a priori information described by geostatistical algorithms with Bayesian inverse problem theory. The extended Metropolis algorithm can be used to sample the a posteriori probability density of highly nonlinear inverse problems, such as full-waveform inversion. Sequential Gibbs sampling is a method that allows efficient sampling of a priori probability densities described by geostatistical algorithms based on either two-point (e.g., Gaussian) or multiple-point statistics. We outline the theoretical framework for a full-waveform inversion strategy that integrates the extended Metropolis algorithm with sequential Gibbs sampling such that arbitrary complex geostatistically defined a priori information can be included. At the same time we show how temporally and/or spatially correlated data uncertainties can be taken into account during the inversion. The suggested inversion strategy is tested on synthetic tomographic crosshole ground-penetrating radar full-waveform data using multiple-point-based a priori information. This is, to our knowledge, the first example of obtaining a posteriori realizations of a full-waveform inverse problem. Benefits of the proposed methodology compared with deterministic inversion approaches include: (1) The a posteriori model variability reflects the states of information provided by the data uncertainties and a priori information, which provides a means of obtaining resolution analysis. (2) Based on a posteriori realizations, complicated statistical questions can be answered, such as the probability of connectivity across a layer. (3) Complex a priori information can be included through geostatistical algorithms. These benefits, however, require more computing resources than traditional methods do. Moreover, an adequate knowledge of data uncertainties and a priori information is required to obtain meaningful uncertainty estimates. The latter may be a key challenge when considering field experiments, which will not be addressed here.

INTRODUCTION

Albert Tarantola was one of the pioneers of seismic full-waveform inversion (see Tarantola, 1984, 1986, 1988). Using a steepest descent algorithm, he obtained the update gradient in each iteration by correlating a forward-propagated wavefield with the residual wavefield propagated backward in time from the receiver positions. This approach has later been referred to as the adjoint method (Talagrand and Courtier, 1987). The first numerical tests based on finite-difference simulations of the seismic signal showed promising results (Gauthier et al., 1986). Since then, several full-waveform inversion algorithms, based on Tarantola's pioneering work, have been developed and applied to seismic data (e.g., Mora, 1987; Crase et al., 1990; Pica et al., 1990; Djikpéssé and Tarantola, 1999).

Ground-penetrating radar (GPR) crosshole tomography is a popular method used to obtain tomographic images of near-surface geological structures and geophysical parameters. The crosshole GPR experiment involves a transmitting radar antenna (20 MHz–1 GHz; see Reynolds, 1997) lowered into a borehole and a receiving antenna placed in an adjacent borehole. The boreholes are typically separated by a distance of 5 m–20 m and are 5 m–100 m deep (e.g., Ernst et al., 2007a; Looms et al., 2008). An antenna is kept fixed in one borehole, while the other antenna is moved between multiple locations in the opposite borehole. The fixed antenna is moved to a new position and the procedure is repeated. At each combination of antennae positions a signal is transmitted between the antennae. In this way an arbitrary dense tomographic data set that covers the interborehole region can be obtained (Peterson, 2001).

Manuscript received by the Editor 16 May 2011; revised manuscript received 25 November 2011; published online 24 February 2012.
[1]Technical University of Denmark, Department of Informatics and Mathematical Modeling, Lyngby, Denmark. E-mail: kcor@imm.dtu.dk; tmeha@imm.dtu.dk; kmos@imm.dtu.dk

Ernst et al. (2007a, 2007b) introduced and applied the adjoint-based optimization algorithm for inversion of tomographic crosshole GPR full-waveform data. They demonstrate that subwavelength features can be resolved when the full-waveform information of the GPR signal is accounted for during the inversion, which was a considerable improvement, compared with traditional ray-based inversion strategies. The inversion scheme of Ernst et al. (2007b) was later adopted for inversion of seismic waveform data by Belina et al. (2009). Recently, the algorithm of Ernst et al. (2007b) has been improved such that it exploits the full vector field of the electromagnetic wave propagation and allows for arbitrary antennae geometry (Meles et al., 2010). Klotzsche et al. (2010) demonstrate an application of the improved code to crosshole GPR waveform data acquired in a gravel aquifer. Lately, Meles et al. (2011) demonstrated that the adjoint-based method of Ernst et al. (2007b) becomes less sensitive to the starting model when the inversion is conditioned to the long wavelengths of the signal in the initial part of the inversion and higher frequencies are gradually incorporated.

The adjoint-based approach is desirable because, in relatively few iterations, it is able to obtain a model of geophysical parameters that minimize a misfit (i.e., objective) function between the observed and modeled waveforms. However, this inversion strategy has some limitations. First, the method is based on linearization of the inverse problem and, therefore, it cannot be guaranteed that the global minimum is found (e.g., Tarantola, 2005). Second, the convergence criterion is chosen somehow subjectively, which may adversely result in noise propagating into the model estimate (e.g., Ernst et al., 2007b). Finally, the method is based on a linear assumption of the forward relation and is limited to Gaussian data uncertainty and a priori information, which results in a Gaussian approximation of the a posteriori uncertainty estimate (Tarantola, 1984; Pratt and Worthington, 1990). Here we propose an algorithm that naturally deals with these limitations.

In a Bayesian formulation, the solution to the inverse problem is given as an a posteriori probability density (Tarantola and Valette, 1982). The a posteriori probability is based on the independent states of information provided by the data (related to the model parameters through a physical law), an associated data uncertainty model (that takes into account data noise and data simulation inadequacies), and the a priori information on the model. The combined states of information contained in the a posteriori probability density are reflected in the model variability of the a posteriori sample. Hence, data uncertainties will not cause artifacts in the solution, but rather influence the degree of a posteriori model variability if the nature of the uncertainties is appropriately accounted for in the uncertainty model. Resolution analysis is naturally obtained from the a posteriori model variability. One may simply be interested in calculating the covariance of the a posteriori model parameters, but more sophisticated questions, such as the probability of geological connectivity or the residence time of a fluid, may also be answered by the a posteriori statistics (Mosegaard, 1998).

The extended Metropolis algorithm (Mosegaard and Tarantola, 1995) can be used to sample the a posteriori probability density, even for highly nonlinear inverse problems. This algorithm does not need a closed form mathematical expression of the a priori information, but a "black box" algorithm that is able to sample the a priori probability density is sufficient. Hansen, et al., (2008; 2012) suggested a method that provides a means of controlling the perturbation step size and efficiency when sampling a priori information defined through any geostatistical algorithms that is based on sequential simulation (e.g. Gomez-Hernandez and Journel, 1993). This method is referred to as sequential Gibbs sampling (Hansen et al., 2012). They demonstrate that this method could serve as a black box algorithm in the extended Metropolis algorithm. Sequential Gibbs sampling can be used to sample a priori models based on either relatively simple two-point based statistical models, such as Gaussian-based a priori models, or more complex multiple-point-based statistical models. This provides a means of using complex a priori statistical models that allow reproduction of geologically plausible structures, such as channels and tortuosity. Such complex patterns (i.e., spatial autocorrelation) can be learned from so-called training images and reproduced by simulation algorithms based on multiple-point statistics (e.g., Strebelle, 2002). Multiple-point algorithms offer the flexibility of simulating realizations with high entropy (e.g., Gaussian distributions) as well as low entropy (i.e., a few facies) structures or a combination of both (Journel and Zhang, 2006). See for example Remy et al. (2008) for different examples of such complex statistical models. In this way, the extended Metropolis algorithm becomes very flexible with regard to the choice of a priori model.

Since his seminal work on the full-waveform inverse problem, Albert Tarantola had the vision that realistic a priori information for inversion could be learned from a large collection of "training images" of the subsurface (Mosegaard, 2011). In this paper we demonstrate how this is made possible by using the extended Metropolis algorithm in conjunction with a priori information defined by a geostatistical algorithm using sequential Gibbs sampling. Initially, the theoretical background for this inversion strategy is outlined. Subsequently, the theory is applied to a tomographic crosshole GPR full-waveform inverse problem in which the a priori information is inferred (i.e., learned) from a training image and realized through the geostatistical algorithm Single Normal equation SIMulation (Snesim) (Strebelle, 2002). Full-waveform data traces are often contaminated by noise and are, in addition, subject to uncertainties related to inadequacies in the data simulation algorithm. These components of data uncertainty do often exhibit some degree of temporal correlation. In this study, we consider a Gaussian-distributed data uncertainty component with a temporal autocorrelation. We show how this uncertainty is accounted for in the inversion through the data uncertainty model.

Tarantola (2005) was a proponent of the movie strategy, in which "movies" of multiple realizations from the a priori and a posteriori probability densities are compared, to understand the additional information provided by the data compared with the a priori information. Features that occur frequently in the a posteriori movie are regarded as well resolved. We show that the movie strategy provides a means of obtaining resolution analysis of the full-waveform inverse problem. Moreover, we demonstrate how a posteriori realizations can be used to quantify the probability of lithological connectivity. The term resolution is, in this paper, used such that high a posteriori model variability refers to low resolution and vice versa (Mosegaard, 1998). To our knowledge this is the first example of sampling the a posteriori probability density of a tomographic full-waveform inverse problem. Resolution analysis of the problem reveals that the combined states of information from the full-waveform data and the a priori information provides a high-resolution subsurface image, even in the case of considerably sparse data

coverage. Finally, we demonstrate how the geostatistically formulated a priori information serves as a guide in the initial burn-in part of the inversion procedure. In this way, convergence of the suggested full-waveform inversion algorithm becomes independent of the initial model.

The present method is completely explicit with regard to the data uncertainty model and a priori information. Establishment of an adequate a priori information and data uncertainty model (comprising data noise and modeling inadequacies) demands effort from the user to obtain trustworthy a posteriori probability density. These issues are far from trivial when considering field experiments and are beyond the scope of this work.

METHODOLOGY

Consider that the subsurface can be represented by a discrete set of model parameters, **m** (referred to as the model) and that a data set, **d**, of indirect observations of the model parameters is provided. The model parameters describe some physical properties of the subsurface that influence the data. Hence, the forward relation between the model and the data can be expressed as the relation (e.g., Tarantola, 2005)

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

in which g is a linear or nonlinear mapping operator that often relies on a physical law. In this particular study the model represents a tomographic image of dielectric permittivity of the subsurface material, and data are waveforms of the vertical electrical component of electromagnetic waves propagated across the model between two boreholes. The forward relation in equation 1 is given as a finite-difference time-domain (FDTD) solution of Maxwell's equations (Ernst et al., 2007b). However, any numerical wave propagation modeling strategy (for GPR or seismic signals) can be applied. The inverse problem is to infer information about the model parameters on the basis of a set of data (and their uncertainties), a priori information about the model, and the forward relation between the model and the data.

In a Bayesian formulation the solution to the inverse problem is given as an a posteriori probability density, which can be formulated as (e.g., Tarantola, 2005)

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

where c is a normalization constant, $\rho_{\mathrm{M}}(\mathbf{m})$ is the a priori probability density, and $L(\mathbf{m})$ is the likelihood function. $\rho_{\mathrm{M}}(\mathbf{m})$ describes the probability that the model satisfies the a priori information. $L(\mathbf{m})$ describes how well the modeled (i.e., simulated) data explain the observed data, given a statistical description of the data noise and modeling inadequacies (from here on referred to as the data uncertainty model). Hence, the a posteriori probability density describes the resulting state of information on the model parameters provided by the independent states of information given by the data (related to the model through the forward relation) and an a priori state of information on the model parameters. Hence, an adequate specification of the data uncertainty model (through the likelihood function) as well as the a priori information are crucial in order to ensure a correct a posteriori state of information (i.e., solution of the inverse problem).

Sampling the a posteriori probability density

A highly nonlinear inverse problem refers to the case in which the a priori probability density is far from being Gaussian or the likelihood function is highly non-Gaussian (typically) due to the nonlinear forward relation between model and data. According to equation 2 the product of these non-Gaussian probability densities signifies a highly non-Gaussian a posteriori probability density. In the case of full-waveform inversion, the forward relation is expected to be highly nonlinear. Moreover, realistic a priori information, as we introduce here, is typically far from being Gaussian.

The extended Metropolis algorithm is a versatile tool which, in particular, is useful to sample the a posteriori probability density of nonlinear inverse problems using arbitrarily complex a priori information. This algorithm is convenient in that it does not need an explicit expression of the a priori probability density. A black box algorithm that is able to perform a random walk in the a priori probability density is sufficient (Mosegaard and Tarantola, 1995). In this study we use sequential Gibbs sampling, which will be described below, as the black box algorithm for sampling the a priori probability density.

The extended Metropolis Algorithm consists of two randomized steps:

1) Exploration: one proposes a candidate model, $\mathbf{m}_{\mathrm{propose}}$, which is a perturbation of a current model, $\mathbf{m}_{\mathrm{current}}$, and at the same time is a realization of the a priori probability density.
2) Exploitation: one decides if the proposed model should be accepted or rejected. The proposed model is accepted with the Metropolis acceptance probability (referred to as the Metropolis rule)

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

where $L(\mathbf{m}_{\mathrm{propose}})/L(\mathbf{m}_{\mathrm{current}})$ is the ratio between the likelihood evaluated in the proposed and the current model, respectively. If accepted, the proposed model becomes the current model and is a realization of the a posteriori probability density. Otherwise the proposed model is rejected and the current model counts again. Thus, in each iteration, the sample size of the a posteriori probability density increases.

The exploration step constitutes the strategy by which proposed models are drawn from the a priori probability density. For small exploration steps the proposed model will be relatively highly correlated with the current model, compared with large exploration steps. Thus, according to equation 3, small exploration steps will result in a high acceptance probability and vice versa. Recall that each evaluation of the likelihood function involves a computationally expensive (FDTD) forward calculation. It is, therefore, important to choose an appropriate exploration step size that does not explore the a posteriori probability density too slowly, but on the other hand does not waste too many expensive evaluations of the likelihood that are very unlikely to be accepted. In particular, the exploration strategy becomes very important when dealing with high-dimensional probability densities, since traditional sampling strategies may lead to a very inefficient exploration strategy (Hansen et al., 2008; 2012). Recently, Hansen, et al., (2008; 2012) introduced a flexible sampling strategy to sample high-dimensional a priori probability densities

defined by geostatistical algorithms. This sampling strategy provides a means of using a variable exploration step size when sampling such high-dimensional probability densities. This method is briefly outlined in the following.

Sampling geostatistically defined a priori information

Hansen et al. (2012) suggest a strategy referred to as sequential Gibbs sampling, which is a method that is capable of sampling probability densities defined by geostatistical algorithms based on sequential simulation. Originally, sequential simulation was used to generate realizations of two-point based statistical models, such as Gaussian models (e.g., Gomez-Hernandez and Journel, 1993). Two-point statistical models are limited to specify the spatial variability between pairs of data (hence the name two-point statistics) defined by a covariance model. Two-point statistics does not provide enough information to model plausible geological structures, such as channels and tortuosity. This can be overcome by using a model based on multiple-point statistics, as suggested by Guardiano and Srivastava (1993). In multiple-point-based statistical models, the a priori information is learned from a training image. The training image is scanned by a template that jointly considers spatial variability among a number of (more than two) pixel values in the image to obtain a joint probability distribution that holds information about these spatial correlations. The model parameter values are subsequently sequentially simulated from conditional probabilities on the basis of the jointly considered pixel values (e.g., Strebelle, 2002). The algorithm originally suggested by Guardiano and Srivastava (1993) was, however, computationally unfeasible. It originally was not until an efficient way of storing multiple-point statistics in machine memory was proposed that the use of multiple-point based models became computationally feasible (Strebelle, 2002). See e.g., Remy et al. (2008) for numerous examples of the application of sequential simulation from both two-point and multiple-point-based a priori models.

In this study sequential Gibbs sampling serves as the black box algorithm that samples the a priori probability density $\rho_M(\mathbf{m})$, described by a geostatistical simulation algorithm, with a controllable exploration step size. The flow of the sequential Gibbs sampling is as follows:

1) An initial unconditional realization of the a priori probability density, $\mathbf{m}_{\text{current}}$, defined in a 2D regular grid of model parameters is provided.
2) A square subarea, corresponding to model parameters $\mathbf{m}_{\text{subarea}}$, with side-length E_{step}(exploration step size) of the current model $\mathbf{m}_{\text{current}}$ is randomly chosen.
3) A realization of the conditional probability density $\rho_M(\mathbf{m}_{\text{subarea}}|\tilde{\mathbf{m}}_{\text{current}})$ is obtained using sequential simulation. $\tilde{\mathbf{m}}_{\text{current}}$ are the current model parameters outside the subarea. $\rho_M(\mathbf{m})$ is an a priori probability density that may be described by either two- or multiple-point statistics. This step is an application of the Gibbs sampler where sequential simulation is used to efficiently generate a realization from the conditional probability density (hence the name sequential Gibbs sampling). In practice this is performed simply by running the sequential simulation algorithm conditional to the model parameters outside the subarea. In this way a new, perturbed model, $\mathbf{m}_{\text{propose}}$ is obtained.
4) The proposed model becomes the current model and steps 2 and 3 are repeated to obtain multiple realizations of the a priori probability density.

Recall that, when applying sequential Gibbs sampling as an a priori model sampler in the extended Metropolis algorithm, the current model $\mathbf{m}_{\text{current}}$ is reused if the proposed model $\mathbf{m}_{\text{propose}}$ is rejected by the Metropolis rule. Moreover, the resemblance (i.e., correlation) between the current and proposed model, and thus the average Metropolis acceptance probability (cf. the Metropolis rule), can be controlled by the explorations step size E_{step} (i.e., the side length of the resimulated area).

The likelihood function and correlated data uncertainties

GPR or seismic full-waveform data are often contaminated with temporally correlated uncertainties along the individual waveform traces or spatially correlated (i.e., static) errors among data related to certain transmitter (source) or receiver positions. The probabilistic formulation of the inverse problem allows for an arbitrary data uncertainty model and it is, therefore, possible to account for these correlations in the data uncertainties. In the present study we consider that data uncertainties are Gaussian distributed with a temporal correlation along the individual waveform traces, but are uncorrelated among the individual traces. This type of uncertainty influences the state of information on the model parameters provided by the data and is, therefore, accounted for through the likelihood function. This particular likelihood function takes on the following form

$$L(\mathbf{m}) = c \prod_{k=1}^{K} \exp\left[-\frac{1}{2}(g(\mathbf{m})^k - \mathbf{d}_{\text{obs}}^k)^{\text{T}}\mathbf{C}_{\text{D}}^{-1}(g(\mathbf{m})^k - \mathbf{d}_{\text{obs}}^k)\right], \tag{4}$$

where $g(\mathbf{m})^k$ and $\mathbf{d}_{\text{obs}}^k$ are vectors that contain the simulated and observed waveform traces related to the kth transmitter-receiver pair. K is the total number of waveform traces (i.e., transmitter-receiver pairs). The factor c is a normalization constant. The term $\mathbf{C}_{\text{D}}$ is the data covariance matrix that defines the variances and covariances of the data uncertainties. The temporal correlation of the data uncertainties is described by an exponential correlation function (e.g., Goovaerts, 1997)

$$\mathbf{C}_D(i,j) = c\,\exp\left(\frac{-3\,s(i,j)}{a}\right), \tag{5}$$

where c is the sill (i.e., variance) and a is the range (i.e., correlation length) of the uncertainties. The term $s(i,j)$ is the temporal distance between the ith and jth sample point along the waveforms. $\mathbf{C}_{\text{D}}$ is a symmetric $N \times N$ matrix, where N is the number of samples in the individual waveform traces.

Cordua et al. (2009) quantified the influence of static-like errors in crosshole GPR experiments, which are data uncertainties that are spatially correlated among data related to the individual transmitter and receiver positions. This kind of data uncertainty may also be accounted for through the data covariance matrix, but this is not considered here. For a description of how to set up a data covariance matrix that accounts for static (i.e., spatially correlated) errors, see Cordua et al. (2008).

The burn-in period

If the probability that a sampling algorithm at any time enters an infinitesimal neighborhood N_j, that surrounds the model $\boldsymbol{m}_j$, is

equal to the probability that it leaves this neighborhood, the algorithm is said to satisfy microscopic reversibility. It can be shown that if a sampling algorithm satisfies microscopic reversibility, then the a posteriori probability density is the only equilibrium distribution for the algorithm, and the algorithm will converge toward this equilibrium distribution independently of the starting model. This property is satisfied by the extended Metropolis algorithm (see Mosegaard and Sambridge [2002] for a detailed description of the Metropolis algorithm). In practice, the extended Metropolis algorithm has reached the equilibrium distribution when the likelihood values start to fluctuate around a constant level (referred to as the equilibrium level) and the data are fitted within their uncertainties. When this phase is reached the algorithm is said to be burned-in. In case of a nonlinear full-waveform inverse problem the structure of the a posteriori probability density is completely unknown. If this distribution is multimodal, several modes may exist that fit data with the uncertainty. Hence, no guarantee can be given, no matter the choice of sample strategy, that the a posteriori probability density is appropriately represented (i.e., that all modes are represented) by a finite sample size.

The a posteriori probability density is defined over a high-dimensional space and leaves only small areas of significant probability. Therefore, the burn-in period may be long because the algorithm during this period randomly walks across large parts of this high-dimensional space searching for a small area of significant probability. As a consequence, we suggest performing large exploration steps in the initial part of the burn-in period and then gradually decreasing the exploration step size as the algorithm approaches the equilibrium level. In this way, the large-scale structures of the model are relatively quickly established in the initial part of the burn-in period, after which the algorithm gradually establishes the smaller-scale structures. However, microscopic reversibility cannot be guaranteed when the exploration step size is not kept constant while running the algorithm. Therefore, the algorithm is stopped when "apparent" burn-in has been reached and the exploration step size is subsequently set to a constant value.

The algorithm is started in an unconditional realization of the a priori model, which is expected to be far away from volumes of significant a posteriori probability density. We suggest adaptively adjusting the exploration step size during the burn-in period such that the Metropolis acceptance probability (equation 3) controls the exploration step size. This is controlled by updating the exploration step size after every M successive exploration steps of the Metropolis algorithm. The exploration step size update is performed just after the exploitation step (step 1) of the algorithm. Consider that the algorithm is at iteration number K, where K is a multiple of M. Then the exploration step size E_{step}^{i+1} during iteration number $K+1$ to $K+M+1$ (i.e., the next M iterations) is given as

$$E_{\text{step}}^{i+1} = E_{\text{step}}^{i} \frac{P_{\text{average}}}{P_{\text{control}}}, \tag{6}$$

where E_{step}^{i} is the exploration step size during iteration number K-M to K (i.e., the M preceding iterations). P_{average} is the average Metropolis acceptance probability (equation 3) during iteration number K-M to K and P_{control} is a subjectively chosen constant acceptance ratio that the algorithm tries to adhere to by adjusting the exploration step size according to equation 6. For larger values of P_{control}, the exploration step size decreases faster and converges to a lower level, than it does for smaller values. The minimum possible exploration step size involves resimulation of one model parameter, whereas the maximum step size is an (unconditional) simulation of all the model parameters that are statistically independent of the previous model.

A SYNTHETIC CROSSHOLE GPR FULL-WAVEFORM INVERSE PROBLEM

The methodology outlined above is tested on a synthetic tomographic crosshole GPR full-waveform data set. Wavefield simulations of the GPR signals (i.e., the forward relation) are obtained using FDTD calculations of Maxwell's equations in transverse electric mode (Ernst et al., 2007b). This FDTD method provides grid-based time-domain calculations of the electromagnetic wavefield propagation. The transmitting and receiving antennae are simulated as vertically orientated dipole-type antennae and are aligned parallel with the vertical boreholes. Transmitted and received signals concern the vertical component of the electrical field (Holliger and Bergmann, 2002). The FDTD algorithm by Ernst et al. (2007b) yields second-order accuracy in both time and space, and performs the calculations in 2D Cartesian coordinates. The edges of the FDTD grid are surrounded by a generalized perfectly matched layer (GPML) to absorb artificial boundary reflections.

The a priori information of the inverse problem is provided by the Snesim algorithm. Snesim is a fast geostatistical simulation algorithm that produces realizations (conditional or unconditional to point data) from a high-dimensional probability density that contains the spatial relations (i.e., patterns) learned from a training image for a relatively low number of categorical values (Strebelle, 2002). Figure 1 shows a training image that mimics a matrix of unconsolidated sand with embedded channels of gravel situated in an unsaturated environment. The geological information contained in the training image may have been obtained from outcrops in a nearby gravel pit and/or a natural cliff. The training image applied here does not necessarily represent a realistic near-surface environment, but is rather used to demonstrate the principle that geologically realistic features, such as complex channel structures, can be

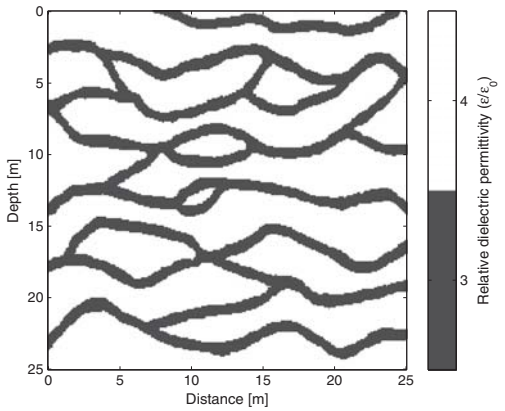


Figure 1. Training image containing geological information of the environment at which the crosshole GPR experiment is conducted. This information is used as a priori information in the waveform inversion.

represented by a multiple-point-based a priori model in the inversion procedure. Electromagnetic signals in the near-surface sediments are sensitive to the dielectric permittivity, the electrical conductivity, and the magnetic permeability of the materials. In this study we limit ourselves to considering only the influence of the dielectric permittivity, which is primarily governing the phase velocity of the signal. The relative dielectric permittivity of the sand and the gravel channel is set to $\varepsilon_r = 4.0$ (0.150 m/ns) and $\varepsilon_r = 3.0$ (0.173 m/ns), respectively (e.g., Ernst et al., 2006). The relative dielectric permittivity is given as $\varepsilon_r = \varepsilon/\varepsilon_0$, where ε is the absolute dielectric permittivity and ε_0 is the dielectric permittivity of free space.

Figure 2 shows the synthetic reference model to be considered and is, at the same time, an unconditional realization of the training image obtained using Snesim. The electrical conductivity is set to a constant value of 3 mS/m and in the following it is assumed known. Near-surface materials are considered nonmagnetic and the magnetic permeability is set to the magnetic permeability of free space (e.g., Davis and Annan, 1989).

A full-waveform synthetic data set is calculated using the FDTD algorithm. A Ricker wavelet with a central frequency of 100 MHz is used as the source pulse. The source pulse is assumed known during the inversion. The recorded synthetic GPR full-waveform data are the vertical component of the electrical field. The experiment is conditioned by data from four transmitters (two in each borehole) at depths of 3 m and 9 m. The receivers are equidistantly distributed in the two boreholes with a separation distance of 1.5 m (see Figure 2). Data acquired with a transmitter-receiver angle larger than 45° from horizontal are omitted since, in practice, these data are violated by effects of wave guiding in the boreholes (e.g., Peterson, 2001) and travel paths between the antenna tips instead of the center of the antennae (Irving and Knight, 2005). These effects are, among several other sources of uncertainty, a result of inadequate forward modeling which has to be seriously considered in field experiments. Such effects either have to be handled through a refined forward modeling approach, or accounted for in the likelihood function through a statistical description of the data uncertainties imposed by both data noise and modeling inadequacies. See the discussion for a further treatment of these issues. The data geometry leads to a total of 20 recorded waveform traces. The resulting transmitter-receiver positions are connected with dotted lines in Figure 3 on top of the grid of the 6240 unknown model parameters.

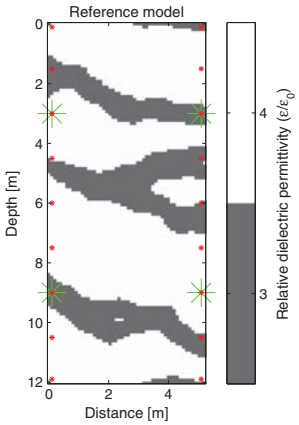


Figure 2. Synthetic reference model. Green asterisks show transmitter positions and the red dots show receiver positions.

Gaussian-distributed data uncertainties with a temporal autocorrelation described by the exponential correlation function in equation 5 are added to the waveform data. The temporal correlation length a (i.e., the range) is set to 12.7 ns. Figure 4 shows the five waveform traces related to the uppermost transmitter position in the left borehole (see Figure 2), which are related to the transmitter-receiver pairs marked by the numbers 1 to 5 in Figure 3. The noise-free simulated waveforms are plotted as dotted blue curves and the uncertain waveforms (noisy waveforms) are plotted as red curves. The 20 uncertain waveform traces are used as observed data in this study and have an average signal-to-noise ratio of 13.6. The uncertainty imposed on the "noise-free" data mimic the total contribution of data noise and modeling inadequacies. In the next section, full-waveform inversion will be performed on the uncertain waveform data with a priori information based on a geostatistical model inferred from the training image in Figure 1.

RESULTS

Burn-in

In the present example the algorithm is started in a realization of the multiple-point a priori model learned from the training image, unconditional to any information from the data. In this way the starting model is independent of data and relies only on the a priori information. The initial exploration step size has a side length of $E_{\text{step}} = 12$ m, which corresponds to the maximum dimension of the model size. Hence, the exploration step size cannot exceed this side length and at this point the algorithm produces statistically

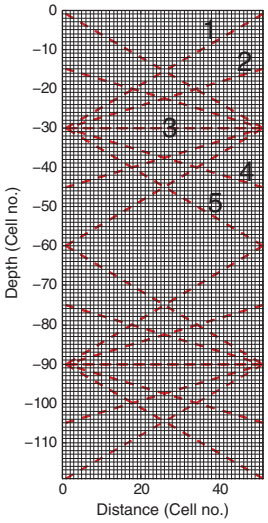


Figure 3. The transmitter-receiver positions connected by dotted lines on top of the 2D grid of ($120 \times 52 = 6240$) unknown model parameters. The lines marked by the numbers 1 to 5 are related to the waveform data shown in Figure 4.

independent realizations. The control acceptance, P_{control}, is set to 10% to obtain a relatively large exploration step size during the burn-in period. The exploration step size is evaluated after each 20th iteration (i.e., $M = 20$) according to equation 6. The evolution of the adaptive exploration step size during the first 1000 iterations is seen in Figure 5. It is observed that the exploration step size is constant and high in the very first part, after which it gradually decreases and stabilizes at a constant level of approximately $E_{\text{step}} = 2$ m. The associated development of the model is demonstrated in Figure 6. It is seen that the large-scale structures are very quickly brought into place, whereupon only fine-scale features of the model are accepted by the Metropolis rule (equation 3). After approximately 17,000 iterations the likelihood values start to fluctuate around an equilibrium level and the data residuals resemble a normal distribution with approximately the same standard deviation as the distribution of the noise. At this stage it is assumed that the algorithm has reached burn-in and produces representative realizations of the a posteriori probability density.

A posteriori statistics

The last model accepted by the Metropolis rule in the burn-in period is used as the starting model when the Metropolis algorithm is subsequently restarted with a constant exploration step size of 1 m. The results obtained in this study demonstrate that the equilibrium level does not change after the Metropolis algorithm is restarted with a constant exploration step size. In addition the data residuals of this period resemble a normal distribution with a standard deviation of $2.17 \cdot 10^{-4}$. As a comparison the standard deviation of the normal distributed noise is $2.24 \cdot 10^{-4}$, which demonstrates that the data are fitted within the data uncertainties.

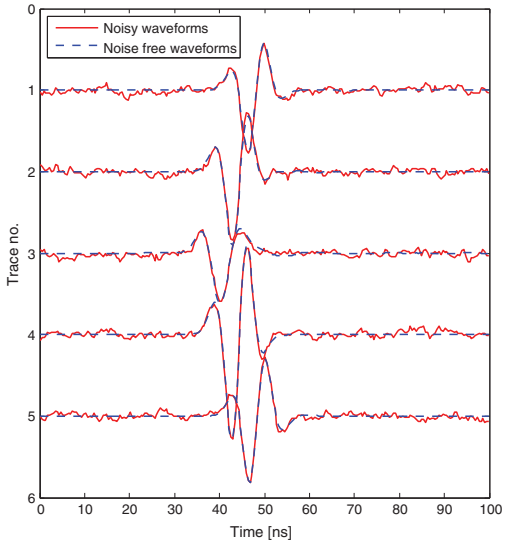


Figure 4. Five waveform traces with and without added noise. The signal-to-noise ratios of the plotted waveform traces are (from the top) 15.4, 20.8, 8.8, 41.3, and 22.2, respectively. The waveforms are related to the transmitter-receiver pairs marked by the numbers 1–5 in Figure 3.

Hence, this shows that the algorithm has reached burn-in and that the adaptive exploration step size may be appropriate when the step size converges toward a constant level (which is the case in our study). This burn-in strategy may serve as an approximate way of determining which exploration step size should be used to obtain a certain average acceptance probability. See Gelman et al. (1996) for an investigation on the choice of acceptance probability.

After 300,000 iterations the algorithm was stopped. During the sample period the algorithm had an average acceptance probability of 40%. Figure 7 shows the autocorrelation between the first model after burn-in and its correlation to the next150,000 models obtained from the a posteriori probability density. These models are not statistically independent because any proposed model in the Metropolis algorithm is a perturbation of a current model or when a proposed model is rejected the current model counts again. Therefore, the autocorrelation analysis of the a posteriori sample shows some correlation length between successive models. In the example shown in Figure 7, statistical independence is obtained after approximately 5800 iterations. This point is approximated as the point at which the autocorrelation curve intercepts the average level of the correlation curve after it has converged to a constant level. The average is shown as a dotted line in Figure 7 and is calculated as the average correlation coefficient between iteration 20,000 and 300,000. A similar analysis is performed on 10 models picked at different iteration numbers, equally distributed across the a posteriori sample, which gives an average of 6745 iterations of separation to obtain statistically independent realizations from the a posteriori sample.

In a probabilistic formulation, the solution to the inverse problem is not a single model estimate, but a sample of multiple model realizations drawn from the a posteriori probability density. Each realization is a tomographic image of the subsurface. Displaying multiple images together corresponds to a movie. The strategy of displaying and studying the solution to the inverse problem using such movies is referred to as the movie strategy (Tarantola, 2005). Roughly speaking, the a priori probability density is filtered by the likelihood function that results in the a posteriori probability density. Hence, displaying a "movie" of a priori realizations together with a movie of a posteriori realizations helps one to understand the

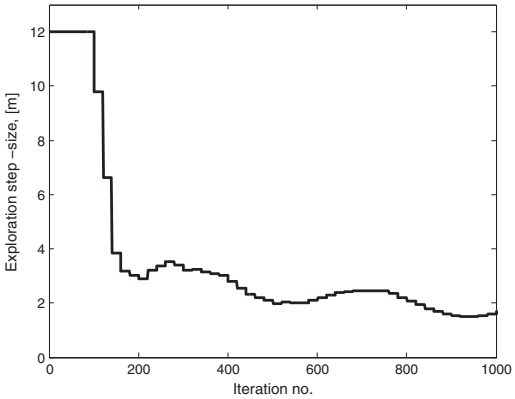


Figure 5. Evolution of the adaptive exploration step size during the first 1000 iterations of the burn-in period.

characteristics of the a priori information imposed on the inverse problem and to understand the state of information provided by the data (and their uncertainties). In particular, when the a priori information is provided through a black box algorithm, and no closed-form mathematical expression of the a priori probability density exists, an a priori movie may be important to understanding whether the state of information provided by the a priori probability density is commensurate with the a priori expectation of the user. See Koren et al. (1991) for a seminal example of using the movie strategy for a seismic inverse problem.

Figure 8 shows eight statistically independent realizations drawn from the a priori probability density (i.e., an a priori movie). This movie shows the state of information provided by the a priori probability density inferred from the training image by the Snesim algorithm. This movie shows a reproduction of the channel structures inferred from the training image, but with no resemblance between the location of the channels in individual realizations, as these models are unconditioned by any data. Figure 9 shows eight statistically independent realizations from the a posteriori probability density. The result clearly demonstrates a high degree of resemblance between the individual a posteriori realizations as a result of the conditioning to the full-waveform data. The high resemblance reveals that the data provides a high resolution of the inverse problem, despite sparse data coverage (see Figure 3). Moreover, the individual realizations only slightly deviate from the reference model (see Figure 2), which confirms the good resolution provided by the full-waveform data, their uncertainties, and the training image. The a posteriori realizations in Figure 9 show some isolated small-scale features, which are not seen in the training image. These noncontinuous effects are caused by the Snesim simulation technique because the algorithm occasionally reduces the number of conditioning data events (i.e., pixel values) to avoid singularities during calculation of the conditional probabilities. A total of 45 statistically independent realizations are obtained from the 300,000 a posteriori realizations based on the model autocorrelation analysis (see Figure 7). These realizations can be used to ask higher-order statistical questions such as what the probability of connectivity is between a channel observed in the left borehole and a channel observed in the right borehole. Figure 9d, 9f, and 9h shows some examples of missing connectivity between the boreholes marked by red circles. For example it is found that in five out of the 45 realizations there is no connection between the points A and B as marked in Figure 9a. Hence, this gives an approximate a posteriori probability of connectivity between points A and B of $(45-5)/45 = 89\%$.

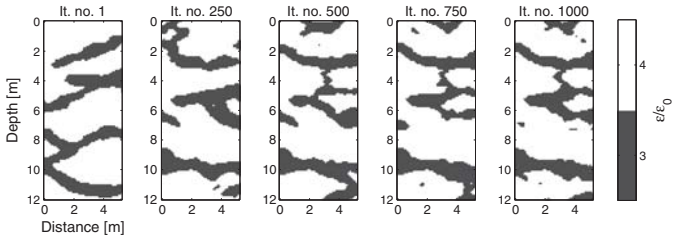


Figure 6. Development of the model during the first 1000 iterations. In this period, the relatively large-scale structures of the model are established. Compare with the reference model in Figure 2. Iteration no. 1 (It. no. 1) is the starting model, which is an unconditional realization of the multiple-point-based a priori model.

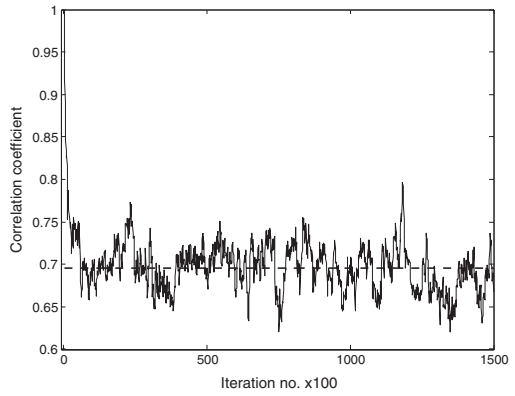


Figure 7. Autocorrelation analysis of the a posteriori sample to determine the number of iterations needed to obtain statistically independent realizations.

Figure 10 shows the mean and variance calculated from the a posteriori sample. It should be noted that the mean model is no longer a realization of the a posteriori probability density, but is only a statistical representation of the sample. The mean tells, in this particular case, the relative a posteriori probability of the presence of a channel at a certain position in the subsurface. The variance reveals that the uncertainty of the spatial location of the channels increases toward the edges of the channels and declines significantly when moving away from the edges.

Waveform data associated with the 45 statistically independent realizations from the a posteriori probability density are calculated for the receiver related to the uppermost transmitter position in the left borehole (see transmitter-receiver pairs marked by the numbers 1–5 in Figure 3). This data variability associated with the model a posteriori variability is plotted in Figure 11a (blue curves) together with the observed data (red curves). The (a posteriori) simulated waveforms show a high degree of similarity and appear in the plot almost as a single fat curve, but are in fact composed of 45 independent waveform curves. Hence, the a posteriori data variability demonstrates that the model a posteriori variability is only associated with very little variability in the waveforms. Moreover, Figure 11a shows that the simulated waveforms fit the observed data very well. Figure 11b shows the data residuals (blue curves) of the simulated waveform data together with the uncertainty component of the observed data (red curves). This plot reveals that the residuals (i.e., misfits) approximately fluctuate around the uncertainty component and resemble the noise statistics (variance and temporal autocorrelation) satisfactorily.

DISCUSSION

Hansen et al. (2006) demonstrated that linear inverse Gaussian theory and simple kriging (i.e., two-point statistics) can be merged

into one single theory. The method of sequential Gibbs sampling used with the extended Metropolis algorithm has provided a further step toward bridging probabilistic inverse problem theory and the field of geostatistics, even for highly nonlinear and non-Gaussian inverse problems (Hansen et al., 2008; 2012). Hansen et al., (2009; 2012) discussed how the complexity of inverse problems (i.e., the time needed to obtain an a posteriori sample) is reduced when a statistical a priori model with some degree of spatial correlation between the model parameters is considered. They showed that the effective dimension of the solution space of the inverse problem, using two-point-based a priori models, is considerably decreased when the a priori expected range of spatial correlation is increased. On the other hand, when no spatial correlation was considered, the high-dimensional inverse problem became unsolvable. Further, it was seen that whether the model parameters take values from a set of real or binary numbers had insignificant influence on the effective dimension compared with the degree of correlation between the model parameters. Hence, we consider the reduction of the effective dimension due to a priori defined spatial autocorrelation to be one of the keystones that makes Monte Carlo inversion of a computationally hard, full-waveform inverse problem feasible, even with thousands of (here 6240) model parameters.

In this study we wish to demonstrate that we are able to freely choose multiple-point-based a priori information, defined by, e.g., the Snesim algorithm, for Monte Carlo-based inversion. We considered a model with categorical parameters only, but this choice of using a multiple-point-based a priori model was a first step away from using simplified a priori models like the Gaussian model. We believe that such models are often chosen out of mathematical convenience rather than for the sake of geophysically and geologically based a priori expectations.

On the basis of the studies of Hansen et al. (2009), we believe that the solution space of the full-waveform inverse problem (i.e., the complexity of the inverse problem), considered in this study, is primarily reduced due to the multiple-point-based spatial correlations, and not significantly due to the binary model parameters. Accordingly, full-waveform inversion may also be tractable when making use of two-point-based statistical a priori models with both continuous and categorical model parameter values, as long as some degree of spatial correlation can be considered a priori. This is also confirmed by our preliminary studies on this topic. The above discussion is encouraging with regard to the possibility of using a priori models based on either two-point statistics, multiple-point statistics, or a combination of both, as long as the chosen a priori model imposes some degree of spatial autocorrelation of the model parameters (see Journel and Zhang [2007] and further discussions at the end of this section). This provides a flexible tool for defining an appropriate a priori model that actually captures our a priori expectations.

The suggested inversion strategy is very general and may be equally applicable for tomographic inversion of any kind of data (e.g., GPR, seismic, x-ray, or electroencephalography data) or for reflection seismic inversion. In the example presented here the algorithm only inverts for the dielectric permittivity, whereas the electrical conductivity is kept fixed. The method could, however, be extended to invert for both parameters by introducing a step, just before the exploration step, that randomly chooses in which of the two fields the exploration should be performed.

A pseudo-full-waveform inversion approach for tomographic GPR data was proposed by Gloaguen et al. (2007). In their approach, multiple model realizations were obtained using a stochastic ray-based inversion strategy. Full-waveform simulations were subsequently calculated in these multiple models. Models related to waveform data that showed the best fit to the observed data were regarded as estimates of the waveform inversion. However, their approach does not guarantee a data fit within the uncertainties and the accepted models are not realizations from an a posteriori probability density function.

In this study the sequential Gibbs sampler is applied such that a continuous block of model parameters are resimulated in each step. Irving and Singha (2010) also used a type of sequential Gibbs sampling, but resimulated a subset of model parameters scattered randomly across the model. Note that sequential Gibbs sampling will correctly sample the a priori probability density function regardless

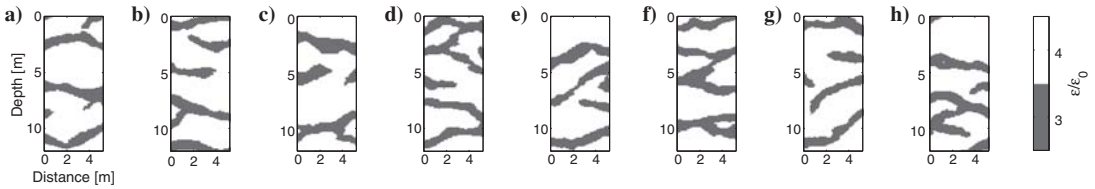


Figure 8. Eight statistically independent realizations of the a priori probability density (a priori movie).

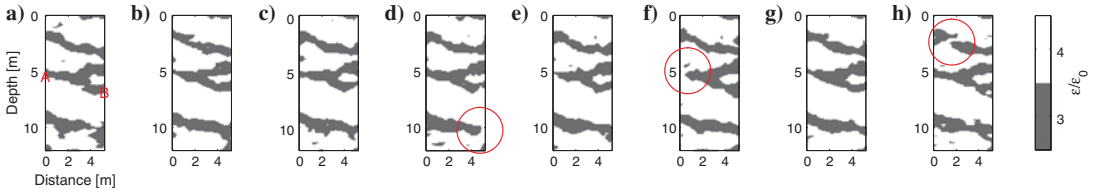


Figure 9. Eight statistically independent realizations of the a posteriori probability density (a posteriori movie).

of how that subset of model parameters to be resimulated is chosen (Hansen et al., 2012). Whether the continuous or the scattered exploration strategy should be used is still to be investigated, as such a choice may affect the computational efficiency when the sequential Gibbs sampler is used with the Metropolis algorithm. To understand exploration strategies of high-dimensional spaces, more analytical approaches are required.

Ray-based inversion experiments on the same scale as the setup used in this study typically involve in the order of 700 to 1600 transmitter-receiver pairs to obtain a reasonable resolution (e.g., Tronicke and Holliger, 2005; Looms et al., 2008; Nielsen et al., 2010). In this study a high degree of resolution was obtained with as few as 20 transmitter–receiver pairs. However, in field experiments modeling inadequacies may lead to considerably more uncertainty in the data than considered in this study, which in turn leads to a lower resolution. In crosshole GPR full-waveform inversion the long spatial wavelengths of the model are typically used as a starting model. This model is obtained through ray-based inversion of first-arrival traveltimes and amplitudes of the waveform data (Ernst et al., 2007a; Meles et al., 2010). The data set considered in our study is expected to be too sparse to provide any useful information for a ray-based starting model. We, therefore, choose to start the full-waveform inversion in an unconditional realization of the a priori probability density, and burn-in is obtained anyway. Hence, the role of the a priori model is more than simply finding a posteriori model realizations that jointly honor data and the a priori model. It turns out that the use of a consistent a priori model acts as a guide in the burn-in process that allows the initial model to be far away from the true solution. The successful convergence from the data-independent starting model observed in the present study may be explained as a reduction of the complexity of the problem through the informative a priori information defined through the geostatistical algorithm (see Hansen et al., 2009; 2012). This encouraging observation suggests that future effort should be toward incorporating complex statistical a priori information into (adjoint) optimization based inversion approaches.

The most commonly used method for full-waveform inversion today is based on adjoint methods, as suggested by Tarantola (1984). This approach has some limitations: (1) Uncertainty estimates may theoretically be obtained through an a posteriori covariance, but only for a linear approximation of the forward relation limited to a Gaussian description of the data uncertainty and a priori model (Tarantola, 1984). (2) The method is based on simple Gaussian a priori information (if any at all); and (3) it relies on a subjective convergence criterion that may adversely result in data uncertainties propagating into the model estimate. The method we propose overcomes many of the limitations of using the adjoint-based approach. (1) It allows for arbitrary data geometry and density. (2) Complex a priori inversion can be included using any geostatistical

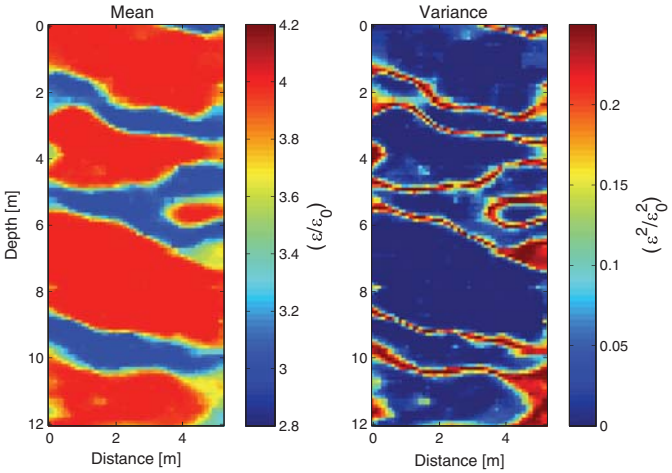


Figure 10. The mean and variance of the model parameters calculated on the basis of the a posteriori sample.

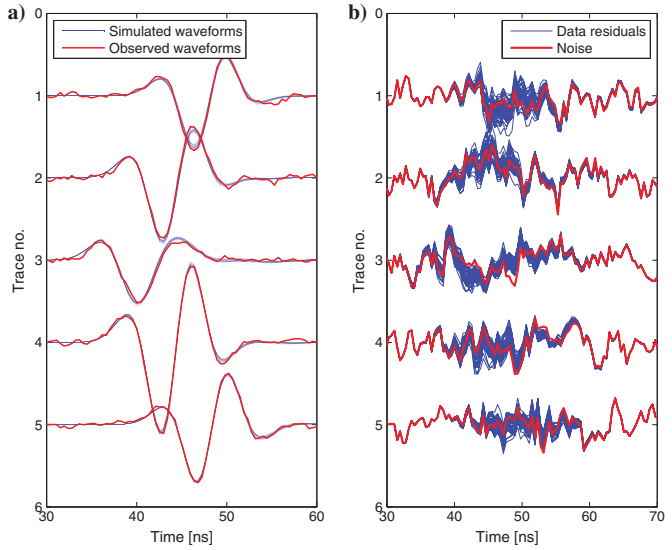


Figure 11. (a) A posteriori data variability. Blue curves: Simulated waveforms calculated for 45 statistically independent a posteriori models. Red curves: The observed waveform data. (b) Data residuals. Blue curves: The data residuals for 45 statistically independent a posteriori models. Red curves: The uncertainty component of the observed data. The waveforms are associated with the transmitter-receiver pairs indicated by the numbers 1–5 in Figure 3.

algorithm that can be used for sequential simulation or any other probabilistically defined a priori information (see Mosegaard and Tarantola, 1995). (3) The full uncertainty of the inverse problem can be quantified by analyzing the a posteriori sample. (4) There are no complications related to the incorporation of refined and complex forward algorithms. (5) Temporally and spatially correlated data uncertainties can be accounted for either through the data covariance matrix or any other mathematical description of the uncertainties (Cordua et al., 2008, 2009). (6) Finally, if the correct data uncertainty model is known there is no risk of uncertainties propagation into a single model estimate, but data uncertainties are instead reflected in the a posteriori model variability.

The adjoint-based approach has some benefits that the presented method lacks: In the case of dense data coverage, full-waveform inversion based on the adjoint-based approach may not need any a priori model at all, as the data themselves are sufficient to allow inference of a solution (Tarantola, 1984). The adjoint-based method has been successfully tested on field data (Ernst et al., 2007a; Klotzsche et al., 2010). In addition the suggested probabilistic inversion strategy needs substantially more (computationally expensive) forward calculations than does the traditional adjoint-based approach.

In the synthetic test performed here, each iteration involves (1) a model perturbation using Snesim which takes approximately 0.5 s, and (2) a forward calculation of the four transmitter positions, which is run in parallel (the number of parallel computations equals the number of transmitter positions). Each forward calculation takes approximately 12.5 s on a standard desktop computer with an Intel Core i7 processor. Thus, the total computation time of 300,000 iterations becomes approximately 45 days. As a comparison, the adjoint-based method needs approximately 20 iterations to converge to a solution estimate (Ernst et al., 2007a), which involves 20 FDTD forward calculations (for comparison with observed data), 20 backward (in time) calculations (for update direction), and 20 forward calculations (for step length) (Ernst et al., 2007b). Hence, obtaining one estimate using the adjoint-based method is approximately 5000 times faster than the time needed to obtain 45 independent realizations from the a posteriori probability density using the Monte Carlo method. However, the adjoint-based method and the Monte Carlo strategy are not directly comparable. The adjoint method is an optimization method that searches for one (in some sense) optimal solution. Additionally, the adjoint method assumes the full-waveform inverse problem to be a global optimization problem and, therefore, the method is always in risk of getting trapped in a local minimum. The Monte Carlo strategy, on the other hand, aims at characterizing the a posteriori probability density. Moreover, it can mathematically be shown that the Metropolis algorithm will converge toward the correct equilibrium distribution independently of the complexity of the data uncertainty and the a priori model (e.g., Mosegaard, 1998). Roughly speaking, the adjoint-based method serves as a fast way of obtaining an approximate maximum a posteriori estimate based on a limited data uncertainty and a priori model. The strategy suggested here could subsequently be started in this estimate to obtain a characterization of the correct a posteriori probability density (i.e., resolution analysis) based on a realistic data uncertainty and a priori model and the full nonlinearity of the forward relation. The significant time expense of the Monte Carlo strategy may in the future be mitigated through parallelization of the individual FDTD calculation on, for example, clusters or graphical processing units (GPU) and by running several algorithms in parallel.

In this study the suggested algorithm was applied on a synthetic test case with a simple reference model. Perfect knowledge about the data uncertainty model, the a priori information, and the source pulse was assumed. Moreover, the electrical conductivity was assumed known. Hence, the robustness of the algorithm has to be further tested by considering more complex synthetic models of both the dielectric permittivity and electrical conductivity and larger data sets. According to equation 2, knowledge about the statistical properties of the data uncertainties and the a priori information evenly influences the a posteriori probability density function. Hence, to ensure a trustworthy a posteriori probability density the user needs to specify a realistic statistical description of the data uncertainties and the a priori information. This may, however, be a very challenging task when dealing with field experiments. On the other hand, if one does not choose the a priori information the inversion algorithm will implicitly make this choice. For example, least-squares and adjoint-based inversion algorithms implicitly consider a Gaussian a priori and data uncertainty model. In the method that we propose here, one is free to choose a Gaussian description of the a priori and data uncertainty model, but the user has the flexibility to choose other more complex descriptions.

In field experiments the list of sources of uncertainties involves both a background noise component and (typically) a major component related to modeling errors (i.e., discrepancies between observed data and simulated data due to shortcomings of the forward relation), such as: (1) source pulse uncertainty, (2) 2D assumption, (3) local antennae coupling effects, (4) effects of discretization, (5) small-scale near borehole heterogeneities, (6) numerical attenuation, (7) media dispersion, (8) point-dipole antenna assumption, (9) effects of high angle travel paths due to wave guiding and antennae spread, (10) unknown antennae positions.

Some of the above mentioned modeling inadequacies can be accounted for through more sophisticated FDTD algorithms (see e.g., Bergmann et al., 1998; Holliger and Bergmann, 2002; Ernst et al., 2006; Irving and Knight, 2006). Alternatively, these modeling uncertainties should be accounted for through the data uncertainty model (although it results in higher a posteriori model variability). Forward modeling inadequacies may lead to a combination of correlated and uncorrelated data uncertainties that may be accounted for through a data covariance matrix (e.g., Cordua et al., 2008; 2009) or a more complex description. See e.g., Peterson (2001), Cordua et al. (2008, 2009), and Irving and Knight (2005) for descriptions and quantifications of some of these uncertainties.

Statistics of the background noise component could be calculated in the signal recorded before the first arriving signal at the receiver. In this period only noise is recorded at the receiver. This may be obtained by performing an experimental covariance analysis of this part of the signal and then subsequently fitting an analytic covariance model to the experimental covariance. The analytic covariance model can then be used to set up an appropriate likelihood function that accounts for the inferred noise component. For more details on how to determine an experimental covariance model (i.e., semivariogram) and how to fit an analytic covariance model see e.g., Journel and Huijbregts (1978) or Goovaerts (1997).

Ernst et al. (2007a) suggested a method based on deconvolution for source pulse estimation, which may also be applied together with our inversion strategy. Then uncertainty related to this process

needs to be quantified or a Bayesian formulation of the source pulse determination could be an integral part of the inversion (see Buland and Omre [2003] for an example of Bayesian wavelet estimation).

Modeling error statistics related to the 2D assumption may be determined through simulation. Consider that a 3D a priori model is established: Forward simulations from, say, 100 realizations of the 3D a priori model could be compared with forward simulations obtained in the associated 2D profile in the transmitter-receiver plane of the 3D model. By subtracting 3D and 2D simulations an estimate of the modeling error statistics associated with the 2D assumptions could be obtained. In the same way modeling error statistics related to other inadequacies in the forward algorithms may be simulated by comparing the results from simple algorithms with more adequate algorithms. In this way adequate computationally expensive forward calculations can (in principle) be substituted with an appropriate data uncertainty model and a faster approximate forward simulation.

While the use of training images to describe a priori information allows complex a priori information to be quantified, the establishment of a priori information through a training image for field data experiments may be challenging. The training image represents a concept of the patterns expected a priori independently of the observed data. These expectations may be based on previous studies in the area based on, for example, kriging and inversion conditional to soft (e.g., seismic) and hard (borehole) data and/or studies of outcrops (e.g., Zhang, 2008). To compromise with the distinct facies (i.e., low entropy) seen in the training image applied here, the a priori model could constitute a combination of information from a training image and a high entropy (e.g., two-point statistically) based a priori model. In any case, we suggest that the movie strategy should be used to ensure that the algorithm or the mathematical expression (or a combination of both) that describes the a priori information also reflects the a priori expectations of the user. See Zhang (2008) for a description of how to convert geological information into training images that can be used with multiple-point statistical simulation algorithms. Journel and Zhang (2006) demonstrated that even for a training image with combined high- and low-entropy information that only approximately captures the complex low entropy structures of the true model, a better conditional modeling result was obtained than for a pure high entropy Gaussian a priori model.

To apply the proposed full-waveform inversion strategy to field data, the major challenge concerning a way of handling modeling uncertainties (probably through a combination of a statistical description and a more adequate modeling algorithm) and to obtaining realistic a priori information that involves training images (from previous studies) still needs to be addressed. Finally, challenges regarding mitigating the computational expense of the Monte Carlo strategy are left for future research.

CONCLUSION

We have outlined the theoretical background for a Monte Carlo-based full-waveform inversion strategy based on the extended Metropolis algorithm in conjunction with complex geostatistical based a priori information. The use of geostatistical algorithms for the description of a priori information can be accomplished in an efficient way through the method of sequential Gibbs sampling, which allows for inclusion of a priori information described by any geostatistical algorithm based on sequential simulation. This, in turn, provides a means of using a priori information described by both two-point and multiple-point statistical a priori models. Inclusion of such statistical a priori information reduces the complexity of the inverse problem, which is a keystone in the feasibility of performing Monte Carlo sampling of the computationally hard full-waveform inverse problem. We have demonstrated the potential of this inversion strategy by sampling the a posteriori probability density of a tomographic full-waveform inverse problem using complex a priori information inferred from a training image using the geostatistical algorithm Snesim. The methodology provides a means of evaluating the a posteriori uncertainty, which is not provided using traditional adjoint-based optimization strategies for full-waveform inversion. However, it should be noted that, if the goal is a single inverse estimate based on pure Gaussian statistics, the adjoint-based optimization approach for inversion of full-waveform data is computationally considerably faster than the suggested inversion strategy. Establishment of adequate (e.g., a multiple-point-based) a priori information and a data uncertainty model for field data are critical to obtain meaningful a posteriori uncertainty estimates. Moreover, the major computational expenses of the Monte Carlo strategy have to be mitigated, which are all challenging tasks that need future research.

ACKNOWLEDGMENTS

We thank DONG E&P, Denmark, for financial support. We which to thank James Irving and two anonymous reviewers who provided thoughtful and valuable comments that helped improve the manuscript.

REFERENCES

Belina, F. A., J. R. Ernst, and K. Holliger, 2009, Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches: Journal of Applied Geophysics, **68**, no. 1, 85–94, doi: 10.1016/j.jappgeo.2008.10.012.

Bergmann, T., J. O. A. Robertsson, and K. Holliger, 1998, Finite-difference modeling of electromagnetic wave propagation in dispersive and attenuating media: Geophysics, **63**, 856–867, doi: 10.1190/1.1444396.

Buland, A., and H. Omre, 2003, Bayesian wavelet estimation from seismic and well data: Geophysics, **68**, 2000–2009, doi: 10.1190/1.1635053.

Cordua, K. S., M. C. Looms, and L. Nielsen, 2008, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data: Vadose Zone Journal, **7**, 263–271, doi: 10.2136/vzj2007.0008.

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**, no. 1, 71–84, doi: 10.1016/j.jappgeo.2008.12.002.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538, doi: 10.1190/1.1442864.

Davis, J. L., and A. P. Annan, 1989, Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy: Geophysical Prospecting, **37**, 531–551, doi: 10.1111/GPR.1989.37.issue-5.

Djikpéssé, H. A., and A. Tarantola, 1999, Multiparameter l_1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035, doi: 10.1190/1.1444611.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007a, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64, doi: 10.1190/1.2761848.

Ernst, J. R., K. Holliger, H. R. Maurer, and A. G. Green, 2006, Realistic FDTD modelling of borehole georadar antenna radiation: Methodology and application: Near Surface Geophysics, 19–31, doi: 10.3997/1873-0604.2005028.

Ernst, J. R., H. Maurer, A. G. Green, and K. Holliger, 2007b, Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations: IEEE Transactions on Geoscience and Remote Sensing, **45**, 2807–2828, doi: 10.1109/TGRS.2007.901048.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403, doi: 10.1190/1.1442188.
Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient Metropolis jumping rules: *in* J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds., Bayesian statistics, **5**, Oxford University Press599–607.
Gloaguen, E., B. Giroux, D. Marcotte, and R. Dimitrakopoulos, 2007, Pseudo-full-waveform inversion of GPR data using stochastic tomography: Geophysics, **72**, J43–J51, doi: 10.1190/1.2755929.
Gomez-Hernandez, J., and A. Journel, 1993, Joint sequential simulation of multi-Gaussian fields: *in* A. Soares, ed., Geostatistics Troia, **92**, Proceedings of the 4th International Geostatics Congress, Kluwer Academic Publishers, 85–94.
Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford Univsity Press.
Guardiano, F., and R. Srivastava, 1993, Multivariate geostatistics: Beyond bivariate moments, *in* A. Soares, ed., Geostatistics Tróia '92, v. 1: Kluwer, 133–144.
Hansen, T. M., A. G. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian theory and geostatistics: Geophysics, **71**, R101–R111, doi: 10.1190/1.2345195.
Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2012, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling: Computational Geosciences, doi: 10.1007/s/10596-011-9271-1.
Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2008, Using geostatistics to describe complex a priori information for inverse problems, *in* J. M. Ortiz, and X. Emery, eds., Geostatistics 2008: Chile, vol. **1**.
Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2009, Reducing complexity of inverse problems using geostatistical priors: IAMG 2009: Stanford, Ca.
Holliger, K., and T. Bergmann, 2002, Numerical modeling of borehole georadar data: Geophysics, **67**, 1249–1257, doi: 10.1190/1.1500387.
Irving, J. D., and R. J. Knight, 2005, Effect of antennas on velocity estimates obtained from crosshole GPR data: Geophysics, **70**, no. 5, K39–K42, doi: 10.1190/1.2049349.
Irving, J. D., and R. J. Knight, 2006, Numerical simulation of antenna transmission and reception for crosshole ground-penetrating radar: Geophysics, **71**, no. 2, K37–K45, doi: 10.1190/1.2187768.
Irving, J. D., and K. Singha, 2010, Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities: Water Resources Research, **46**, no. 11, W11514, doi: 10.1029/2009WR008340.
Journel, A., and T. Zhang, 2006, The necessity of a multiple-point prior model: Mathematical Geology, **38**, no. 5, 591–610, doi: 10.1007/s11004-006-9031-2.
Journel, A. G., and C. J. Huijbregts, 1978, Mining geostatistics: Academic Press.
Klotzsche, A., J. van der Kruk, G. A. Meles, J. A. Doetsch, H. Maurer, and N. Linde, 2010, Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland: Near Surface Geophysics, **8**, 635–649, doi: 10.3997/1873-0604.2010054.
Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299, doi: 10.1029/91JB02278.
Looms, M. C., K. H. Jensen, A. Binley, and L. Nielsen, 2008, Monitoring unsaturated flow and transport using cross-borehole geophysical methods: Vadose Zone Journal, **7**, no. 1, 227–237, doi: 10.2136/vzj2006.0129.
Meles, G., S. Greenhalgh, J. van der Kruk, A. Green, and H. Maurer, 2011, Taming the non-linearity problem in GPR full-waveform inversion for high contrast media: Journal of Applied Geophysics, **73**, no. 2, 174–186, doi: 10.1016/j.jappgeo.2011.01.001.
Meles, G. A., J. van der Kruk, S. A. Greenhalgh, J. R. Ernst, H. Maurer, and A. G. Green, 2010, A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data: IEEE Transactions on Geoscience and Remote Sensing, **48**, 3391–3407, doi: 10.1109/TGRS.2010.2046670.
Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228, doi: 10.1190/1.1442384.
Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: Inverse Problems, **14**, 405–426, doi: 10.1088/0266-5611/14/3/004.
Mosegaard, K., 2011, Quest for consistency, symmetry, and simplicity — The legacy of Albert Tarantola: Geophysics, **76**, no. 5, W51–W61, doi: 10.1190/geo2010-0328.1.
Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 431–447, doi: 10.1029/94JB03097.
Mosegaard, K., and M. Sambridge, 2002, Monte Carlo analysis of inverse problems: Inverse Problems, **18**, R29–R54, doi: 10.1088/0266-5611/18/3/201.
Nielsen, L., M. C. Looms, T. M. Hansen, K. S. Cordua, and L. Stemmerik, 2010, Estimation of chalk heterogeneity from stochastic modelling conditioned by crosshole GPR travel times and LOG data: *in* R. Miller, J. Bradford, and K. Holliger, eds., Advances in near-surface seismology and ground-penetrating radar: SEG Geophysical Development Series 15, 379–398.
Peterson, J. E., 2001, Pre-inversion correction and analysis of radar tomographic data: Journal of Environmental and Engineering Geophysics, **6**, no. 1, 1–18, doi: 10.4133/JEEG6.1.1.
Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292, doi: 10.1190/1.1442836.
Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310, doi: 10.1111/GPR.1990.38.issue-3.
Remy, N., A. Boucher, and J. Wu, 2008, Applied geostatistics with SGeMS: A user's guide: Cambridge University Press.
Reynolds, J. M., 1997, An introduction to applied and environmental geophysics: John Wiley and Sons.
Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, **34**, no. 1, 1–21, doi: 10.1023/A:1014009426274.
Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory: Quarterly Journal of the Royal Meteorological Society, **113**, 478, 1311–1328, doi: 10.1256/smsqj.47811.
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.
Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903, doi: 10.1190/1.1442046.
Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Pure and Applied Geophysics, **128**, 1-2, 365–399, doi: 10.1007/BF01772605.
Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society of Industrial and Applied Mathematics.
Tarantola, A., and B. Valette, 1982, Inverse Problems = Quest for Information: Journal of Geophysics, **50**, 159–170.
Tronicke, J., and K. Holliger, 2005, Quantitative integration of hydrogeophysical data: Conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers: Geophysics, **70**, no. 3, H1–H10, doi: 10.1190/1.1925744.
Zhang, T., 2008, Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics: Earth Science Frontiers, **15**, no. 1, 26–35, doi: 10.1016/S1872-5791(08)60016-0.

Appendix A5:

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

Authors:

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

Published in:

Expanded abstract of Society of Exploration Geophysicists' 82th Annual Meeting (SEG2012)

Las Vegas, Nevada, USA

4 – 9 November 2012

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):

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

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):

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

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:

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

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:

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

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

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

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 $k \in \{j \mid H_j^{cur} \neq 0\}$. 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:

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

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.

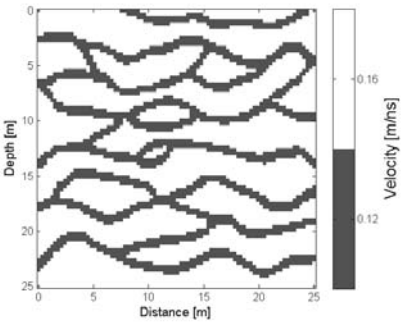


Figure 1. Training image used for obtaining the multiple-point a priori statistics.

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 $H_k^{prior} = 5, \quad k \in \{j \mid H_j^{cur} \neq 0\}$ 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.

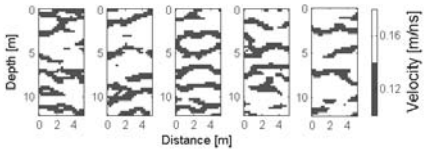


Figure 2. Realizations from the SNESIM a priori model using sequential Gibbs sampling.

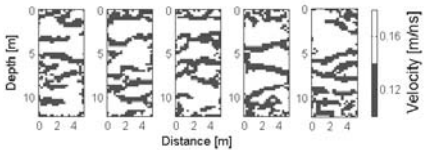


Figure 3. Realizations of the Dirichlet pdf (i.e. FFM-based a priori pdf) using the Metropolis algorithm with a homogenous proposal pdf.

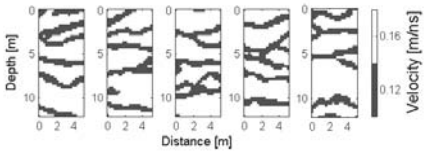


Figure 4. Realizations from the combined SNESIM-FMM-based a priori pdf using the extended Metropolis algorithm in conjunction with sequential Gibbs sampling.

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.

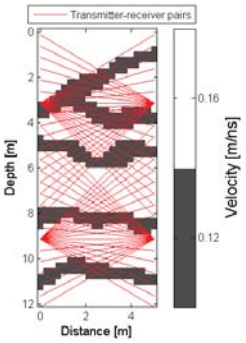


Figure 5. Reference model used for travel time tomography. The red rays give an indication of the data coverage.

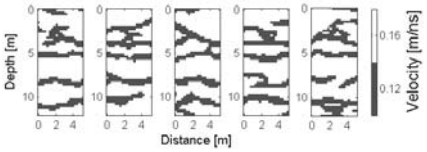


Figure 6. Realizations from the a posteriori pdf with a priori information defined by SNESIM using sequential Gibbs sampling.

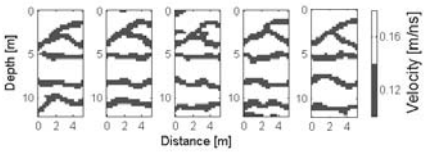


Figure 7. Realizations from the a posteriori pdf based on the combined SNESIM-FMM a priori pdf using sequential Gibbs sampling.

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.

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.

Appendix A6:
Improving the pattern reproducibility of multiple-point-based prior models

Authors:

Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard

Published in:

Submitted to Mathematical Geosciences

Improving the pattern reproducibility of multiple-point-based prior models

Knud S. Cordua[*], Thomas M. Hansen[*], Klaus Mosegaard[*]

* Technical University of Denmark, Center for Energy Resources Engineering, National Space Institute, Electrovej, Building 327, 2800 Kgs. Lyngby, Denmark.

Corresponding author: Knud Skou Cordua, kcor@dtu.dk, fax 45259575, phone +4527260794.

Short title: Improving multiple-point-based prior models

Submitted: 7/3 - 2013

Abstract

Some multiple-point simulation algorithms, such as the SNESIM algorithm, are based on sequential simulation using conditional probabilities obtained from training image statistics. Such algorithms, which we will refer to as pruned sequential sample algorithms, are based on imperfect implementations of sequential simulation because all necessary conditional probability distributions cannot be obtained. Consequently, zero probability multiple-point events may occur during the simulation. The pruned sequential sample algorithms deal with this by reducing (i.e. pruning) the number of conditioning events until a non-zero probability can be established. However, in this way some of the information obtained from the training image is lost, which reduces the reproducibility of the training image patterns in the output realizations. By means of sequential Gibbs sampling, sequential simulation based algorithms can be used as prior models for inverse problems. Sequential Gibbs sampling has been proven to work for perfect implementations of sequential simulation. When the SNESIM algorithm is used with the sequential Gibbs sampler, the pattern reproducibility of the output realizations is deteriorated. In order to mitigate this problem, we suggest combining the sequential Gibbs sampler with the frequency matching method. This combination maintains the efficiency of the sequential Gibbs sampler while seeking to match the multiple-point statistic from the training image and the output realizations. In this way the pattern reproducibility of the SNESIM prior realizations is considerably improved. Finally, a tomographic cross-borehole example is used to demonstrate how the pattern reproducibility of the prior model influences the solution to an inverse problem.

Keywords: Cross-hole tomography, Multiple-point statistics, Training image, Sequential Simulation algorithm, Pattern frequency, Frequency Matching Method.

1. Introduction

In geostatistics and probabilistic inverse problems prior information about the spatial distribution of the model parameters, describing the subsurface properties under study, are important in order to obtain trustworthy predictions of production data or uncertainty quantifications or of the subsurface structures (Journel and Zhang, 2006; Hansen et al., 2013b). Moreover, prior models with high information content may also reduce the effective dimension of the inverse problem and make otherwise computationally heavy inverse problems tractable (Hansen et al., 2009; Cordua et al., 2012). In some cases an unrealistic prior may lead to apparent data uncertainties (Cordua et al., 2009) and may event prevent the observed data to be fitted.

Traditionally, prior models have relied on two-point statistics parameterized by a mean and covariance (e.g. Journel, 1974; Alabert, 1987; Hansen et al., 2008; Cordua et al., 2009). However, such prior models cannot capture realistically geological complex structures such as tortuous channels. In order to overcome this shortcoming, multiple-point statistics was introduced (Guardiano and Srivastava, 1993). The Single Normal Equation SIMulation (SNESIM) algorithm is a computationally efficient way of obtaining prior realizations based on multiple-point statistics "learned" from a training image using sequential simulation (Strebelle, 2002). The SNESIM algorithm has previously been used as prior model for inverse problems conditioned by production data (e.g. Caers, 2006), satellite data (Boucher et al., 2008), tomographic travel time data (Hansen et al., 2008) and full waveform data (Cordua et al., 2012).

Examples of geostatistical multiple point based algorithms not based on sequential simulation are CCSIM (Tahmasebi et al., 2012) and FILTERSIM (Zhang et al., 2006). Another way of expressing complex geological prior information is through object based simulation. In these methods the earth model is represented by a set of objects with geometries and positions defined by probability distributions (Deutsch and Wang, 1996; Skorstad et al., 1999)

Lange et al. (2012) suggested a way of defining a multiple-point-based prior model using a Frequency Matching Method (FMM). This method quantifies the degree of match between the pattern frequency (i.e. multiple-point histogram) of the training image and a realization from this prior model. Lange et al. (2012) quantified the match between frequency distributions using a Chi-square dissimilarity measure. In this way they were able to optimize for a model that, at the same time honored the multiple-point-based prior information from a training image and seismic tomographic travel time data.

In a probabilistic formulation, the solution to an inverse problem is given as a posterior probability distribution over the model parameters, which combines the information provided by a relation between the data and model parameters, the observed data, an associated data uncertainty, and data-independent prior information about the model parameters (see e.g. Tarantola, 2005). For the sake of uncertainty quantification of the model parameters, rather than searching for a set of model parameters with maximum posterior probability, a sample from the posterior distribution has to be obtained. This can be obtained by means of sample algorithms such as the Metropolis algorithms or the rejection sampler (Tarantola, 2005; Irving and Singha, 2010, Mosegaard and Tarantola, 1995; Hansen et al., 2013b). The rejection sampler is only suitable for very low-dimensional problems or problems with very noisy data (Hansen et al., 2013b). The Metropolis algorithm, on the other hand, is appropriate for higher dimensional problems, such as geoscientific inverse problems (Mosegaard and Tarantola, 1995).

The Metropolis algorithm requires an algorithm that is able to perform a random walk in the prior probability distribution with a controllable degree of perturbation between the realizations (i.e. a step-length). An efficient way of obtaining this for sequential simulation based algorithms is through sequential Gibbs sampling (SGS) (Hansen et al., 2008; Hansen et al., 2012). The Probability Perturbation Method (PPM) is another example of a sample strategy that can be used to make a perturbation between realizations from multiple-point based sample

algorithms (Caers, 2006). An example of an algorithm used for perturbations between realizations from two-point statistical prior models is the Gradual Deformation Method (GDM) (Hu, 2000).

The Metropolis algorithm has previously been used in conjunction with Sequential Gibbs Sampling for prior information defined through the SNESIM algorithm to sample the solution of a tomographic full waveform inverse problem (Cordua et al., 2012). Other examples of using a methodology related to sequential Gibbs sampling for sampling of solutions to inverse problems are seen in Hansen et al. (2008, 2013b), Mariethoz et al. (2010a) and Irving and Singha (2010).

Most multiple-point-based sequential simulation algorithms (such as the SNESIM algorithm (Strebelle, 2002), the direct sampling method (Mariethoz et al., 2010b) and other methods that rely on the idea by Guardiano and Srivastava, 1993) are based on an implementation of sequential simulation that simulates from conditional probability distributions based on multiple-point statistics obtained from a training image. We will refer to such algorithms as pruned sequential sampling (PSS) algorithms because these sequential sample algorithms prune the number of conditioning events until a non-zero probability to be used for the simulation can be established. Moreover, we will refer to a perfect implementation of sequential simulation as the situation where a conditional probability distribution conditioned to all previously simulated model parameters is always available. This is not the case for the PSS algorithms because only a limited set of these conditional probability distributions is used by such algorithms. As a consequence, the sequential simulation process may ask for a conditional probability distribution based on a multiple-point event with zero probability according to the training image. This problem is fixed (circumvented) by reducing the number of conditioning events until a multiple-point event with non-zero probability is obtained (for details see Strebelle, 2002). However, this comes at the cost of reducing the multiple-point information propagated from the input training image to the output realizations.

This, in turn, reduces the ability to reproduce the training image patterns in the output realizations.

Hansen et al. (2012) proved that for a prior model based on a perfect implementation of sequential simulation the sequential Gibbs sampler samples this prior model correctly. However, Cordua et al. (2012) noticed that when sequential Gibbs sampling is applied with the SNESIM algorithm, the pattern reproducibility is deteriorated when compared with traditional unconditional SNESIM realizations. This tells us that the multiple-point based information propagated from the training image to the output realizations is further reduced (as compared to generating an unconditional realization using SNESIM) when the SNESIM algorithm is used with the sequential Gibbs sampler. The reason for that will be further addressed in the discussion section. This is undesirable for two reasons: 1) Such a prior model becomes influenced by the number and geometry of the pixels/area used for the local resimulation (i.e. step-length) performed by the sequential Gibbs sampler. Hence, the prior model becomes unsuitable for use with the Metropolis algorithm where the step-length should preferably be chosen for maximum computational efficiency without influencing the information expressed by the prior model. 2) The amount of prior information from the input training image about the subsurface patterns that is propagated to the output realizations is reduced.

In this study we suggest to combine the frequency matching method with the sequential Gibbs sampling using a pruned sequential sample algorithm. In this way we maintain a better match between the multiple-point pattern frequency of the training image and the prior realizations. Moreover, this combination maintains the efficiency of the sequential Gibbs sampler. We define a frequency matching method that uses the Dirichlet probability distribution to quantify the match between the pattern frequency distributions and describe how this can be combined with the sequential Gibbs sampler. The SNESIM algorithm is here used as an example of a pruned sequential sample algorithm. We compare sets of prior

realizations based on: 1) the frequency matching method 2) traditional application of SNESIM without sequential Gibbs sampling, 3) SNESIM based sequential Gibbs sampling, and 4) a combination of the frequency matching method with SNESIM based sequential Gibbs sampling.

The importance of pattern reproducibility in the prior when solving inverse problems is demonstrated through a cross-hole travel time tomography example. The solution to this nonlinear inverse problem is sampled using the Metropolis algorithm with prior information based on both the SNESIM-based sequential Gibbs sampler and the proposed method combining the frequency matching method with the SNESIM-based sequential Gibbs sampler.

2. Methodology

2.1 Probabilistic inverse problem theory

Consider that the subsurface can be parameterized by a discrete set of model parameters $\mathbf{m}$ that represent the subsurface properties under study. In a probabilistic formulation, the solution to the inverse problem is based on combined information about the model parameters. This information is provided by a set of indirect observations $\mathbf{d}$ (e.g. travel time data), data uncertainties $\rho_D(\mathbf{d})$, a theoretical forward problem that describes the relation between the model parameters and data

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

and some data independent prior information on the model parameters $\rho_M(\mathbf{m})$. The forward function g describes a linear or nonlinear relation that relies on a physical law or an emperically based relation. In this study equation 1 is a nonlinear relation that provides a set of travel time data observed at the receiver positions given a 2D velocity field between the

transmitter and receiver positions. The forward relation is based on ray-approximation and is calculated using the Eikonal equation (Zelt and Barton, 1998).

The solution to the inverse problem is given as the posterior probabiltiy distribution over the model parameters (e.g. Tarantola, 2005)

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

where k is a normalization constant and $L(\mathbf{m}) = \rho_D(g(\mathbf{m}))$ is the likelihood function. The prior distribution $\rho_M(\mathbf{m})$ describes the prior information over the model parameters. The likelihood function $L(\mathbf{m})$ describes information over the model parameters provided by the forward problem, the observed data and the uncertainty related to the observed data (i.e. noise).

In this paper we assume that the noise component is Gaussian distributed with zero mean. In this case the likelihood function takes on the form

$$L(\mathbf{m}) = c\exp\left(-\frac{1}{2}\left(g(\mathbf{m}) - \mathbf{d}^{obs}\right)\mathbf{C}_D^{-1}\left(g(\mathbf{m}) - \mathbf{d}^{obs}\right)\right), \qquad (3)$$

where $\mathbf{d}^{obs}$ is a vector of observed data, $\mathbf{C}_D$ is the covariance matrix that describes the (co)variances of the data noise.

2.2 The extended Metropolis algorithm

The extended Metropolis algorithm can be used to sample the posterior probability distribution of a general nonlinear inverse problem as formulated in equation 2. This algorithm requires: 1) A "black box" algorithm that is able to perform a random walk in the prior probability distribution (i.e. produce perturbations between realizations from the prior distributions). 2) An evaluation of the likelihood function for a given set of model parameters. The extended Metropolis algorithm takes the following steps:

1) The exploration step:

A prior sampler proposes a realization, $\mathbf{m}_{propose}$, from the prior probability distribution, where $\mathbf{m}_{propose}$ is a perturbation of a current realization, $\mathbf{m}_{current}$.

2) The exploitation step:

The proposed realization is accepted with probability

$$P_{accept} = \min\left(1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{current})}\right) \quad (4)$$

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

2.3 Sequential Gibbs sampling

In order to make sequential simulation based algorithms appropriate for inverse modeling it is often required that a random walk with a controllable step-length between two prior realizations can be obtained (e.g. the current and proposed models described in the Metropolis algorithm). Sequential Gibbs sampling is such a tool that can be used to control the step-length between realizations obtained from algorithms based on sequential simulation (Hansen et al., 2008; Hansen et al., 2012).

In a perfect implementation of sequential simulation each model parameter is simulated from a conditional probability distribution that is conditioned by all previously simulated model parameters. Hence, all conditional probability distributions from the final joint probability distribution are provided. Consequently, the probability distribution that is sampled in this way can be expressed as

$$p(x_1, x_2, ..., x_N) = p(x_N \mid x_{N-1}, ..., x_2, x_1)\, p(x_{N-1} \mid x_{N-2}, ..., x_2, x_1) ... p(x_2 \mid x_1)\, p(x_1) \quad (5)$$

Using this implementation the sequential simulation will lead to a realization of the same joint distribution independent of the simulation sequence.

Hansen et al. (2012) proved that for a perfect implementation of sequential simulation the sequential Gibbs sampler satisfies detailed balance, which guarantees that once the sequential Gibbs sampler samples the desired distribution it will continue sampling this distribution.

The sequential Gibbs sampler provides a sampling strategy that can be used as prior sampler for the Metropolis algorithm. In this way complex geostatistical sampling algorithms based on sequential Gibbs sampling can be used as prior models for probabilistic inverse problems.

The flow of the sequential Gibbs sampler is:

1) A current realization from the prior distribution 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 from the prior distribution.

The size of the subset of model parameters to be resimulated is chosen subjectively and controls the explorational nature of the extended Metropolis algorithm. For large subsets the exploration step becomes large and the probability of acceptance (see equation 4) 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 e.g. Cordua et al. (2012).

2.3.1 Example

In this section, prior realizations from the pruned sequential sample algorithm SNESIM algorithm and the SNESIM algorithm used in conjunction with sequential Gibbs sampling are compared. The multiple-point statistics used as input for these two prior models is obtained from the training image seen in Fig. 1.

Figure 2 shows 5 traditional unconditional realizations from SNESIM using a template of 7 x 7 pixels. Figure 3 shows realizations using SNESIM-based sequential Gibbs sampling with a resimulation area of 2 m x 2 m (8 pixels x 8 pixels). A comparison of Figs. 2 and 3 reveals that the use of the sequential Gibbs sampling strategy reduces the quality of the pattern reproduction. Moreover, the patterns seen in the output realizations from both of these prior models do not resemble the patterns seen in the input training image. This motivates us to have a look at the frequency matching method, which provides a means of controlling the degree of match between the pattern frequency distribution of the input training image and the output realizations.

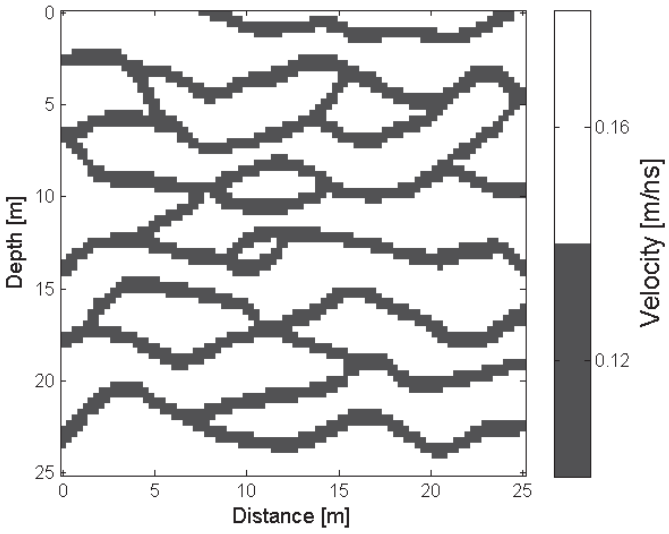


Figure 1. Training image used for obtaining the multiple-point prior statistics.

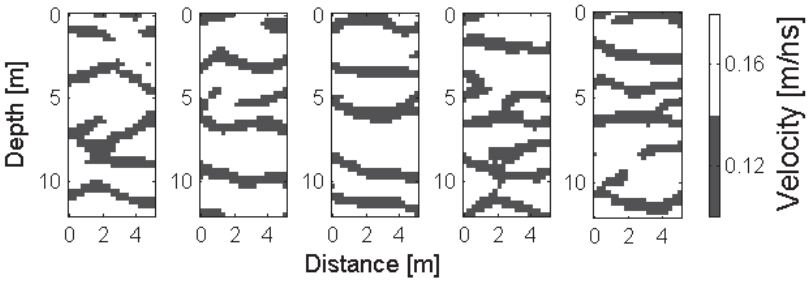


Figure 2. Independent realizations from the SNESIM algorithm without sequential Gibbs sampling.

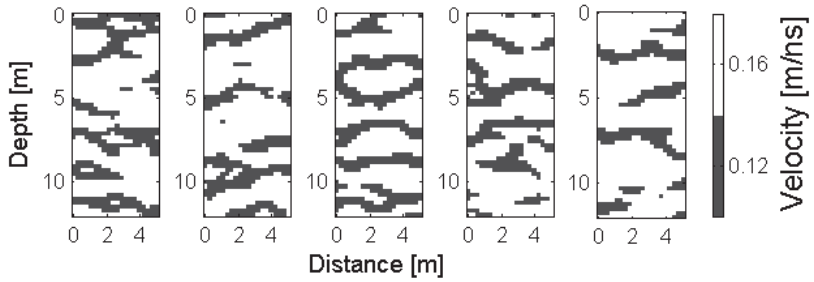


Figure 3. Realizations from the SNESIM prior model using sequential Gibbs sampling.

2.4 The frequency matching method

Multiple-point-based sample algorithms that rely on sequential simulation simulates from local conditional probability distributions (e.g. Strebelle, 2002). These conditional probability distributions are (according to the product rule) expressed by means of a (local) joint probability distribution. The joint distribution is based on multiple-point statistics from the training image, which is obtained by counting the number of times a certain pattern (i.e. multiple-point event) occurs in the training image. The joint probability distribution obtained in this way can be viewed as a pattern frequency distribution (i.e. a multiple-point histogram), which carry the same information as the content of the search tree, as referred to by Strebelle (2002).

In the frequency matching method (Lange et al., 2012) the multiple-point-based prior probability distribution is quantified by measuring the degree of fit between the pattern frequency distribution of the training image and an output realization. In this way it becomes possible to formulate a closed form mathematical expression that quantifies the multiple-point prior probability distribution, which is not possible using a pruned sequential sample algorithm.

Lange et al. (2012) defined a frequency matching method using a Chi-square dissimilarity measure. However, using this formulation does not lead to a prior probability

distribution, which makes it inappropriate to be used in a pure probabilistic formulation of the solution to the inverse problem. Therefore, we chose to defined the frequency matching measure using the Dirichlet probability distribution

$$\rho_{FMM}(\mathbf{m}) = \frac{N^{cur}!}{H_1^{cur}!,\ldots,H_K^{cur}!}\prod_{k=1}^{K}\left(\frac{H_k^{TI}+H_k^{prior}}{N^{TI}+N^{prior}}\right)^{H_k^{cur}}, \quad (6)$$

where H_k^{cur} is the number of counts in the k'th bin of the (unnormalized) histogram obtained from a prior 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:

$$N^{cur} = \sum_{k=1}^{K} H_k^{cur} \quad (7)$$

$$N^{TI} = \sum_{k=1}^{K} H_k^{TI} \quad (8)$$

$$N^{prior} = \sum_{k=1}^{K} H_k^{prior}, \quad (9)$$

where H_k^{prior} is the k'th bin of the prior (unnormalized) histogram, which represents the prior expectation of the pattern histogram before the training image histogram is observed. For high values of H_k^{prior} a mismatch between the frequency distributions of an output realization and the training image has higher probability than for lower values of H_k^{prior}. Notice that the Dirichlet distribution only needs to be evaluated for bins in which H_k^{cur} as non-zero counts. All other bins (with zero counts) do not contribute to the probability. Hence, the histograms

$H_k^{TI} + H_k^{prior}$ becomes sparse, which, in particular, saves memory for large template sizes and/or many categories of the model parameter values.

To our knowledge, this allows, for the first time, the possibility of quantifying the prior probability of a certain model $\mathbf{m}$ given a training image (and the associated pattern frequency distribution). One can sample the prior according to ρ_{FMM} directly using for example using the Metropolis sampler with a uniform proposal distribution. An example of this is shown in Fig. 4. Figure 4 shows realizations from the frequency matching based prior using the Dirichlet distribution in the case choosing the prior histogram to be a homogenous distribution with $H_k^{prior} = 5, \quad k \in \left\{ j \mid H_j^{cur} \neq 0 \right\}$ and using a template size of 3 pixels x 3 pixels.

While the evaluation of equation 6 in itself is computationally very cheap, this is a computationally extremely inefficient way to sample such a multiple point based prior, due the use of the uniform proposal distribution. Instead we propose to combine the frequency matching method, using the definition in equation 6, with a pruned sequential sample algorithm (exemplified here through the SNESIM algorithm) such that the computational efficiency of the sequential Gibbs sampling strategy can be utilized.

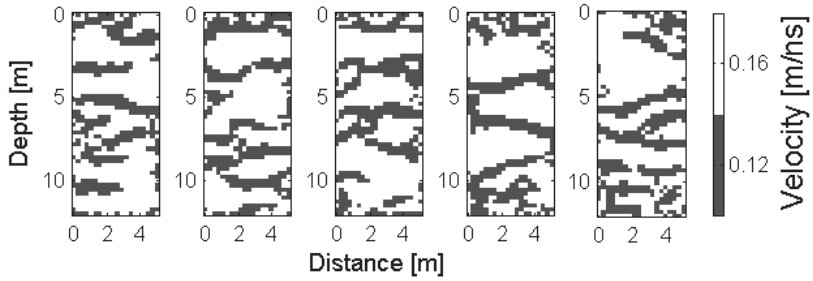


Figure 4. Realizations of the Dirichlet probability distribution obtained using the Metropolis algorithm with a homogenous prior probability distribution.

2.5 Combining the frequency matching method with a PSS algorithm

In this study we suggest to combine (algorithm-driven) PSS-based prior models (such as a SNESIM-based prior model) ρ_{PSS} with the frequency matching method prior based on a mathematic expression as a product of the two

$$\rho_M(\mathbf{m}) = \rho_{PSS}(\mathbf{m})\rho_{FMM}(\mathbf{m}) \quad (10)$$

If the prior pattern frequency distribution H_k^{prior} is assumed homogenous, then for very high values of H_k^{prior} the frequency matching based prior approaches a homogeneous distribution and the combined prior approaches the PSS-based prior. Hence, in principle the combined prior provides a means of controlling the degree of training image pattern match.

The combined prior distribution can efficiently be sampled using the Metropolis algorithm in conjunction with the sequential Gibbs sampler. By substituting $\rho_{FMM}(\mathbf{m})$ with the likelihood function $L(\mathbf{m})$ in equation (2) and (4), realizations from the combined prior in equation 10 can be obtained. The algorithm that samples this combined prior has the following workflow:

1) Generate a current unconditional prior realization $\mathbf{m}_{current}$ from the PSS-based prior $\rho_{PSS}(\mathbf{m})$.

2) Calculate the prior probability of this realization $\rho_{FMM}(\mathbf{m}_{current})$ using the frequency matching method.

3) Use the sequential Gibbs sampler (see section 2.3) to produce a perturbation of the current model in order to obtain a proposed model $\mathbf{m}_{propose}$.

4) Evaluate the propose model using the frequency matching method $\rho_{FMM}(\mathbf{m}_{propose})$.

5) Accept the proposed model with probability

$$P = \min\left[1, \frac{\rho_{FMM}(\mathbf{m}_{propose})}{\rho_{FMM}(\mathbf{m}_{current})}\right] \quad (11)$$

If the proposed model is accepted, $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{current}$, otherwise $\mathbf{m}_{current}$ is accepted (and counted) again. By repeating step 2 – 5 a desirable number of models are accepted. All models accepted by this algorithm will be realizations from the combined prior $\rho_M(\mathbf{m}) = \rho_{PSS}(\mathbf{m})\rho_{FMM}(\mathbf{m})$. Note that, in this way, the prior probability of the PSS-based realization $\rho_{PSS}(\mathbf{m})$ does not need to be evaluated.

The computational efforts needed to evaluate $\rho_{FMM}(\mathbf{m})$ is insignificant compared to performing one iteration of PSS algorithm.

3. Results

Figure 5 shows the results of obtaining realizations from the combined prior using sequential Gibbs sampling in conjunction with the Metropolis algorithm (using the same setting of the sequential Gibbs sampler, SNESIM and frequency matching method as described above). The results demonstrate that realizations from the combined prior model contain the same continuous channel structures as seen in the training image. It is observed that the pattern reproducibility is even better than compared to the traditional unconditional SNESIM realizations not influence by the sequential Gibbs sampler (compare Figs 2 and 5). Realizations using the SNESIM algorithm tends to show chopped off channel structures, while this is not the case using the combined prior model.

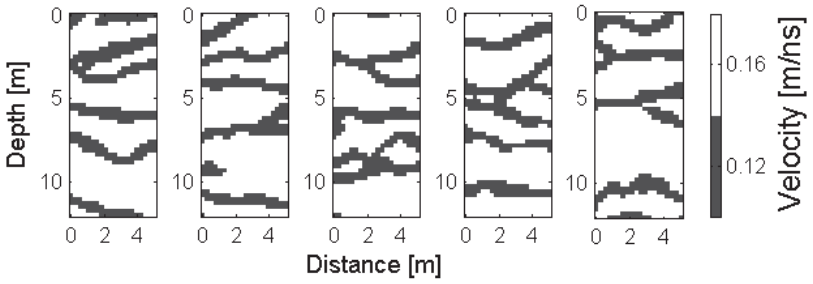


Figure 5. Realizations from the combined SNESIM-FMM-based a priori pdf using the extended Metropolis algorithm in conjunction with sequential Gibbs sampling.

3.1 Influence of the prior on the solution to a tomographic inverse problem

In order to demonstrate how the ability of the prior model to reproduce the patterns of the training image influences the solution to a nonlinear inverse problem, we consider a crosshole ground penetrating radar tomographic inverse problem (see e.g. Cordua et al., 2009). The inverse problem as defined here is formulated in a probabilistic manner as defined by equation 2. Realizations from the posterior distribution are obtained using the Metropolis algorithm.

A synthetic reference model, from which a synthetic data set is obtained, is seen in Fig 6. This model is an unconditional realization of the SNESIM based prior model. A zero mean uncorrelated Gaussian noise component with a standard deviation of 1 ns (~2.7 % of the signal) is added to the data. This noise is appropriately accounted for by the likelihood function as seen in equation 3. The posterior probability distribution of this inverse problem is sampled using the Metropolis algorithm. The result of the inversion using SNESIM-based sequential Gibbs sampling prior is seen in Fig 7 (the associated prior realizations are seen in Fig 3). The posterior realizations related to the inverse problem using the combined prior model are seen in Fig 8. It is clear that the improved combined prior model provides posterior realizations that

resemble the reference model better than when using the SNESIM-based sequential Gibbs sampling prior. Moreover, the variability between the individual realizations becomes smaller when considering the combined prior model. This shows that the improved, and more informative prior, improves the resolution of the solution to the inverse problem. Additionally, the training image pattern reproducibility is improved.

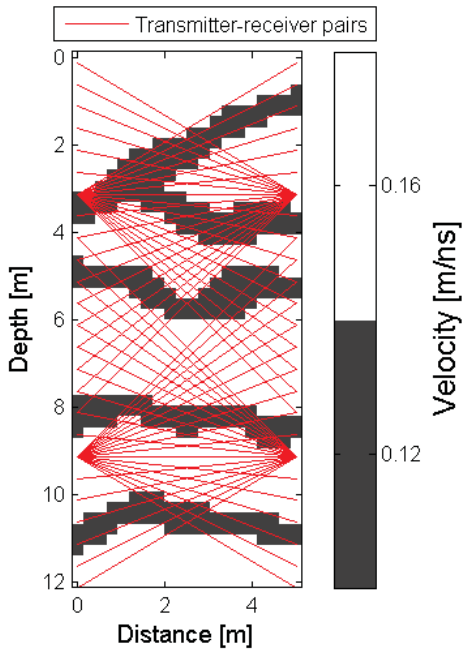


Figure 6. Reference model used for travel time tomography. The red rays give an indication of the data coverage.

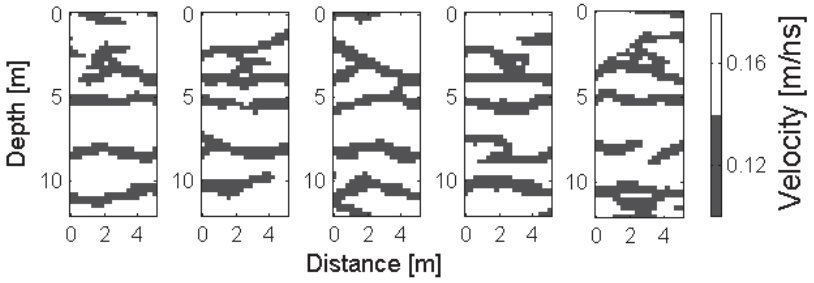


Figure 7. Realizations from the posterior probability distribution with prior information defined by SNESIM-based sequential Gibbs sampling.

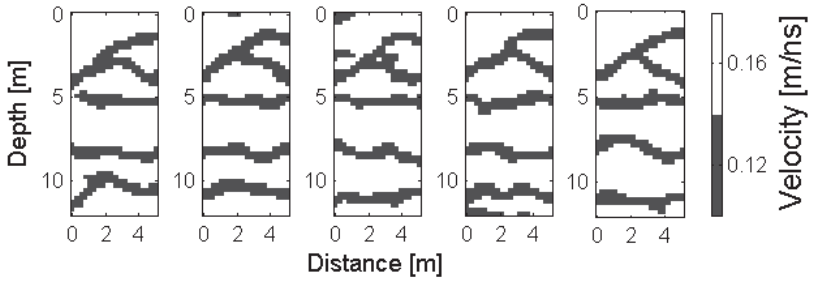


Figure 8. Realizations from the posterior probability distribution based on the prior model combining the frequency matching method and the SNESIM-based sequential Gibbs sampler.

4. Discussion

4.1 PSS based Sequential Gibbs sampling

During the sequential simulation performed by PSS algorithms a zero-probability multiple-point event may occur. This is dealt with by reducing the number of conditioning events until a non-zero multiple-point event is obtained. When the number of conditioning events is reduced, the information contained in the multiple-point statistics that is extracted

from the training image is reduced and, therefore, not propagated to the outcome realizations. This loss of information influences the ability of the algorithm to reproduce the input patterns as seen in the training image.

Figure 3 showed that when sequential Gibbs sampling is used in conjunction with the PSS algorithm SNESIM the reproducibility of the training image patterns in the outcome realizations were worse than compared to full unconditional SNESIM realizations. As the sequential simulation in PSS algorithm proceeds, the likelihood that the algorithm needs to reduce the number of conditioning events will increase because it gradually becomes harder to fit a multiple-point event (as they are obtained from the training image) with non-zero probability. The local resimulation performed by the sequential Gibbs sampler corresponds to being in the last part of the sequential simulation process. Consequently, the number of conditioning events is likely to be reduced, which in turn reduces the number multiple-point information propagated from the training image to the realizations. Visually this will be observed as a reduction in the pattern reproducibility. Moreover, this tells us that the step-length (i.e. the size of the resimulated area) influences the prior information contained in the output realizations.

When using the combined prior, the frequency matching method ensures a better propagation of the multiple-point-based information from the input training image to the output realization by assigning low probability to multiple-point events not observed in the training image. In addition, the prior may become less sensitive to the step-length of the sequential Gibbs sampler.

4.2 Training image patterns vs. prior expectations

As it was seen above, the patterns seen in the training image are not necessarily reproduced in the realizations when using the non-combined prior models. In these cases the

patterns of the training image are not the prior information expressed by the algorithm-driven prior. Hence, the prior information actually expressed by such prior models can only be realized using the movie strategy, which involves looking at a set of prior realizations generated by the actual algorithm that is used as prior model for the inverse problem (see e.g. Cordua et al., 2012).

It has been argued that most all geostatistical simulation methods are algorithm-driven, which means that one must run the algorithm in order to see what kind of spatial variability is actually produced by this prior model. The use of such 'algorithmic prior' adds a layer of complexity between the expert on spatial structures (such as the geologist) and the use of this information as prior information for a modeling problem. If the prior information about the subsurface model expressed by the training image is actually reflecting the prior believes about the subsurface, it is unsatisfying not to maintain this information in the output of the prior model. Moreover, as it is seen above, the pattern reproducibility has an enormous impact on the solution to the inverse problem. The pattern reproducibility might even be more crucial in case of flow simulation in which the scenario of connected vs. disconnected channel results in a significant difference in the forward responses. This study demonstrates the importance of not assuming that the training image in itself represents the prior information, but rather the combination of the training image and the sampling algorithm. Therefore, for any practical use of realistic models of spatial variability it is necessary to develop methods that allow a geological expert to quantify his/her knowledge of spatial variability (in form of e.g. a training image) without needing to be an expert in geostatistics and algorithmic details.

4.3 Maximum posterior models with training image based prior information

Lange et al. (2012) expressed a frequency matching method based prior using a Chi-square dissimilarity distance. For a cross-borehole tomographic inverse problem they were

able to find solution that honored a combination of the likelihood function and this multiple-point-based prior model. The frequency matching method based on the Dirichlet distribution is a purely probabilistic formulation of the prior model and can, therefore, be used to find the maximum posterior model of the probabilistically defined inverse problem

$$\sigma_M(\mathbf{m}) = k\rho_{FMM}(\mathbf{m})L(\mathbf{m}) \quad (12)$$

The advantage of the Dirichlet formulation as compared to the Chi-square dissimilarity measure is that the Dirichlet formulation is a probability distribution where the "weighting" factor H_k^{prior} is based on prior believes and is not found by using a L-curve method that searches to find a balanced weight between the prior and likelihood terms, as is the case for Lange et al. (2012). The use of the frequency matching method based on the Dirichlet distribution is well suited for use with probabilistic inversion methods, both for sampling the posterior distribution and to allow locating the model with maximum posterior probability. Hansen et al (2012) discuss that no methods is yet available to locate the model with maximum posterior probability in the case of a training image based prior, because, until now, no method has existed to quantity the prior probability. This is now the case for the first time using equations 6 and 12 above. An example of finding a model with maximum posterior probability based on a Dirichlet formulated prior is saved for future research.

It should be noted that the influence of the patterns from the training image can be controlled by H_k^{prior}. Therefore, the suggested combined prior can be used to control to what degree the patterns of the prior realization matches the training image patterns. If the influence of the frequency matching method is completely removed (for very high values of H_k^{prior}), then the prior reduces to a prior model based on PSS based sequential Gibbs sampling. On the other hand, for small values of H_k^{prior}, the frequency matching method imposes a high degree of pattern matching on the combined prior model.

4.4 Sampling the frequency matching based prior model

If it was possible to evaluate the probability of the outcome realizations from the PSS algorithm, it would be possible to use this algorithm as proposal for the sample algorithm. In such a case it would be able to sample a pure frequency matching based prior model (instead of the product of the frequency matching based prior and the PSS-based prior model) while maintaining the computational efficiency of the sequential simulation strategy. Such a sample algorithm is obtained by substituting the accept probability in equation 11 with the following

$$P = \min\left[1, \frac{\rho_{FMM}(\mathbf{m}_{propose})\rho_{PSS}(\mathbf{m}_{current})}{\rho_{FMM}(\mathbf{m}_{current})\rho_{PSS}(\mathbf{m}_{propose})}\right], \quad (13)$$

which compensates for the proposal distribution. Using the algorithm in section 2.5 with this accept probability the algorithm will sample $\rho_{FMM}(\mathbf{m})$ instead of $\rho_{PSS}(\mathbf{m})\rho_{FMM}(\mathbf{m})$.

5. Conclusion

We have demonstrated the potential of combining the frequency matching method with a PSS-based sequential Gibbs sampling (here exemplified through the SNESIM algorithm). In this way, the output realization from this combined prior conserves the information from the multiple-point statistics obtained from the training image, which visually appear as an improved reproduction of the pattern from the training image. This reproduction is even better than when compared to traditional unconditional realization from the SNESIM algorithm. The suggested strategy maintains the computationally efficiency of sequential simulation while improving the reproducibility of the multiple-point patterns from the training image.

The suggested combined prior model was used for an inverse problem to demonstrate the influence of the pattern reproducibility to the solution of an inverse problem. We showed that the improved pattern reproducibility improved the resolution when applied

for a tomographic nonlinear inverse problem. This demonstrates the importance of being able to propagate the multiple-point-based prior information from the training image into the output realizations.

Acknowledgements

We thank DONG E&P for financial support.

References

Alabert F (1987) The practice of fast conditional simulations through LU decomposition of the covariance matrix. Mathematical geology, 19 (5): 369 – 386.

Boucher A, Kyriakidis PC, Cronkite-Ratcliff C (2008) Geostatistical Solutions for Super-Resolution Land Cover Mapping. IEEE Transactions on geoscience and remote sensing, 46 (1): 272 – 283.

Caers J, Hoffman T (2006) The probability perturbation method: a new look at Bayesian inverse modeling. Mathematical Geology, 38 (1): 81 -100. doi:10.1007/s11004-005-9005-9

Cordua KS, Hansen TM, Mosegaard K (2012) Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information. Geophysics, 77: H19 – H31.

Cordua KS, Nielsen L, Looms MC, Hansen TM, Binley A (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.

Deutsch CV, Wang L (1996) Hierarchical object-based geostatistical modeling of fluvial reservoirs. Paper SPE 36514.

Guardiano F, Srivastava M (1993) Multivariate geostatistics: beyond bivariate moments. In: Geostatistics - Troia, pp. 133 – 144. Kluwer Academic, Dordrecht.

Hansen TM, Cordua KS, Mosegaard K (2012) Inverse problems with non-trivial priors: Efficient solution through Sequential Gibbs Sampling. Computational Geosciences, 16: 593 – 611. DOI: 10.1007/s10596-011-9271-1.

Hansen TM, Cordua KS, Looms MC, Mosegaard K (2013a) SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 1 — Methodology: Computers and Geosciences, 52: 470 - 480.

Hansen TM, Cordua KS, Looms MC, Mosegaard K (2013b) SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 2 — Application to crosshole GPR tomography: Computers and Geosciences, 52: 481 – 492.

Hansen TM, Mosegaard K, Cordua KS (2008) Using geostatistics to describe complex a priori information for inverse problems. in J. M. Ortiz, and X. Emery, eds., Geostatistics 2008: Chile, vol. 1.

Hansen TM, Mosegaard K, Cordua KS (2009) Reducing complexity of inverse problems using geostatistical priors. IAMG 2009: Stanford, Ca.

Hu LY (2000) Gradual Deformation and Iterative Calibration of Gaussian-Related Stochastic Models. Mathematical geology, 32 (1): 87 – 108.

Irving J, Singha K (2010) Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resources Research, 46: W11514.

Journel AG (1974) Geostatistics for conditional simulation of ore bodies. Econ. Geol., 69: 673 – 687.

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

Lange K, Frydendall J, Cordua KS, Hansen TM, Melnikova Y, Mosegaard K (2012) A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information. Mathematical geosciences, 44: 783 – 803.

Mariethoz G, Renard P, Caers J (2010a) Bayesian inverse problem and optimization with Iterative Spatial Resampling. Water resources research, 46: W11530.

Mariethoz G, Renard P, Straubhaar (2010b) The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46: W11536.

Mosegaard K, Tarantola A (1995) Monte Carlo sampling of solutions to inverse problems. Journal of geophysical research. 100 (B7): 431 – 447.

Skorstad A, Hauge R, Holden L (1999) Well conditioning in a fluvial reservoir model. Math. Geol. 31: 857–872.

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

Tahmasebi P, Hezarkhani A, Sahimi M (2012) Multiple-point geostatistical modeling based on the cross-correlation functions. Computational geoscience, 16: 779 – 797.

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

Zelt C, Barton P (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 (B4): 7187 – 7210.

Zhang T, Switzer P, Journel AG (2006) Filter-based classification of training image patterns for spatial simulation. Math. Geol., 38: 63 – 80.

Appendix A7:

On sequential simulation of incompletely known random fields

Authors:

Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard

Published in:

Submitted to Mathematical Geosciences

On sequential simulation of incompletely known random fields

Knud Skou Cordua[*], Thomas Mejer Hansen[*], Klaus Mosegaard[*]

* Technical University of Denmark, Center for Energy Resources Engineering, National Space Institute, Electrovej, Building 327, 2800 Kgs. Lyngby, Denmark.

Corresponding author: Knud Skou Cordua, kcor@dtu.dk, fax 45259575, phone +4527260794.

Short title: On sequential simulation of incompletely known random fields

Submitted: 7 / 4 - 2013

Abstract

Probabilistic inverse problems and geostatistical modeling rely on and benefit from trustworthy prior information. For such problems, prior information is expressed by a probability distribution over the space of model parameters. In practice, however, prior information is often given as incomplete knowledge extracted from few realizations of the subsurface model (e.g. training images), stemming from earlier data analysis, or provided by specialists. From such information approximate marginal distributions can be derived. In this paper, we discuss how to determine a joint probability distribution (i.e. random field) that is consistent with a limited set of known marginal distributions. Central to our discussion is the use of additional constraints to supplement the marginals when building a well-defined joint distribution, and the effects of these constraints on the resulting distribution. We pay particular attention to the currently most applied method for injection of prior information into inverse problems, namely sequential simulation. Sequential simulation algorithms are capable of sampling joint probability distributions based on a set of marginal distributions. We investigate the implicit assumptions made by such algorithms and show that distributions sampled by such algorithms depend on the sampling sequence, and do not have the desired marginal distributions. We describe how a sequential simulation strategy based on a Markovian joint distribution can avoid these problems. Finally, we address the problem that training images have limited size and therefore, even under an assumption of stationarity, provide limited statistical information about the marginals. For this case, we suggest a formulation that incorporates these statistical uncertainties.

Keywords: Multiple-point Geostatistics, sample algorithms, entropy, information theory, prior model.

1 Introduction

In a probabilistic formulation, the solution to an inverse problem or a geostatistical modeling problem is given by a (posterior) probability distribution, which is both influenced by the observed data and the prior state of information on the model parameters (Tarantola, 2005). Therefore, trustworthy prior probability distributions are important in order to obtain a dependable solution to such problems (Journel and Zhang, 2006; Hansen et al., 2012).

Geostatistical models are useful tools for quantifying geological information. Therefore, such models serve as a means of expressing prior information about an earth model to be inferred (e.g. Cordua et al., 2012). Geostatistical modeling basically rely on two different approaches: (1) the object-based approach in which the geometrical shapes of geobodies are defined by stochastic parameters (e.g. Deutsch and Wang, 1996). (2) The pixel-based approach in which the earth model is characterized by an image where the pixel values are associated with stochastic parameters (Journel and Alabert, 1989). The information provided by a geostatistical model can be expressed by a joint probability distribution defined over a set of stochastic parameters (also referred to as model parameters). Here, our focus will be on the pixel based approach. In this case the individual parameters are associated with a position in space (typically arranged in a regular grid). A joint probability distribution over parameters that, as in our case, refer to spatial locations is known as a random field. However, throughout this paper we will refer to it as a joint probability distribution.

Typically, this joint probability distribution is unknown, except for a limited set of its marginal probability distributions. Assuming stationarity, the marginal distributions can be estimated from a sample model (i.e. a realization from the unknown joint distribution) in form of a training image or an old dataset. The single normal equation simulation (SNESIM) algorithm is an example of a geostatistical algorithm that samples a joint probability distribution based on a set of marginal distributions obtained from a training image (Strebelle, 2002). Another example is to assume the joint distribution to be Gaussian distributed and then determination its parameters using a semivariogram (i.e. covariance model) based on a set of observations (e.g. Journel and Huijbregts, 1978; Looms et al, 2010). The simulation algorithm may use a limited class of functions to represent the marginal distributions in order to save computational time. An example of such an algorithm is sequential Gaussian simulation (SGSIM) using a local neighborhood of conditioning

data (e.g. Journel, 1994). An example of determining a joint probability distribution, which represents a noise model, based on a sample model in form of set of observations, is shown by Mosegaard et al. (1997).

The simplest possible statistics that can be obtained from a sample model (such as a training image) is expressed by a one-dimensional marginal distribution (i.e. a histogram of single-pixel values). However, if information about the spatial dependencies between model parameters has to be contained in the statistics (i.e. two- or multiple-point statistics), higher dimensional marginals have to be estimated (e.g. Lange et al., 2012). An M-dimensional marginal distribution then represents the probability of patterns over certain M-pixel configurations (e.g. Guardiano and Srivastava, 1993; Strebelle, 2002). Markov random fields with higher order interactions have previously been used to characterize large scale complex spatial patterns (Tjelmeland and Besag, 1998). Recently, an approach using a Markov mesh model has been proposed (Stien and Kolbjørnsen, 2011).

In this paper we discuss the general problem of determining a joint probability distribution based on a set of its marginal probability distributions. It is shown that for a limited set of marginal probability distributions this will in general lead to an underdetermined problem. We discuss different assumptions that can be made in order to uniquely define a joint probability distribution given its marginals.

Sequential simulation algorithms are able to sample an unknown joint probability distribution based on a set of marginal distributions. We investigate the assumptions made by such algorithms and in this way obtain explicit mathematical expressions for the joint probability distribution that is sampled by these algorithms. It is demonstrated that if the assumptions are used in an incorrect way the joint distribution will depend on the simulation sequence (i.e. the random path used) and will not be consistent with the marginal distributions. We review the sequential simulation strategy based on a Markovian joint distribution that is consistent with the marginal distributions and is not influenced by the simulation sequence.

Marginal probability distributions estimated from a sample model are subject to uncertainty due to the limited size of the training image or data set used to obtain the sample model. We suggest a formulation of a Markovian joint distribution that takes into account the statistical uncertainty related to the estimation of marginal distributions. Finally, we discuss the

information content (i.e. entropy) of joint probability distributions and how it is influenced by lack of consistency and uncertainty with regard to the marginal distributions.

2 From marginal to joint probability distributions

In this section we investigate the problem of inferring a joint probability distribution based on a limited set of its marginal probability distributions. Consider two marginal probability distributions $p(m_1)$ and $p(m_2)$ of dimension M = 1 from a joint probability distribution $p(m_1, m_2)$ of dimension N = 2. The model parameters are assumed to be categorical variables with K = 3 possible outcomes ($m \in (1,2,3)$). Then, the joint probability distribution has $K^N = 3^2 = 9$ possible outcomes (see Fig 1). The L = 2 corresponding marginal probability distributions have $LK^M = 2 \cdot 3^1 = 6$ possible outcomes.

In order to determine the joint probability distribution depicted in Fig 1, based on the marginal probability distributions, the probabilities of its 9 possible outcomes must be calculated from a total of 6 possible outcomes from the two marginals. This calculation amounts to solving a system of 6 linear equations with 9 unknowns (using p_{11} as short for $p(m_1 = 1, m_2 = 1)$ and p_1 as short for $p(m_1 = 1) = p(m_2 = 1)$, etc.)

$$\begin{aligned}
p_{11} + p_{12} + p_{13} &= p_1 \\
p_{21} + p_{22} + p_{23} &= p_2 \\
p_{31} + p_{32} + p_{33} &= p_3 \\
p_{11} + p_{21} + p_{31} &= p_1 \\
p_{12} + p_{22} + p_{32} &= p_2 \\
p_{13} + p_{23} + p_{33} &= p_3
\end{aligned} \tag{1}$$

subject to the constraint that all $p_i \geq 0$, and an additional equation stating that the probabilities must sum to one:

$$p_{11} + p_{12} + p_{13} + p_{21} + p_{22} + p_{23} + p_{31} + p_{32} + p_{33} = 1. \tag{2}$$

The latter two conditions state that the solution must reside in the unit 9-simplex, part of an 8-dimensional linear subspace of $\mathbf{R}^9$. In matrix notation this system of equations takes on the form

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{31} \\ p_{32} \\ p_{33} \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_1 \\ p_2 \\ p_3 \\ 1 \end{bmatrix} \Leftrightarrow \tag{3}$$

$$\mathbf{A}\mathbf{p}^{\text{joint}} = \mathbf{p}, \tag{4}$$

where $\mathbf{p}^{\text{joint}}$ is the unknown joint probability distribution and $\mathbf{p}$ contains the known marginal distributions. It can be shown that this linear set of equations has an 4-dimensional linear solution space (the matrix $\mathbf{A}$ in equation 4 has rank 5) in $\mathbf{R}^9$. Hence, its solution is non-unique, and the additional constraints are not sufficient to uniquely determine a solution. This can be seen in the following way:

Assume that $\left\{\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \mathbf{n}^{(3)}, \mathbf{n}^{(4)}\right\}$ is an orthonormal basis for the 4-dimensional nullspace of the matrix $\mathbf{A}$ in equation 4. Since the last row of this matrix consists of ones only, the sum of the components of each basis vector must be zero:

$$\sum_i n_i^{(k)} = 0 \tag{5}$$

for $k = 1, 2, 3, 4$. As the basis vectors are orthogonal, all their components have an absolute value smaller than one. Assume that the numerically largest value of all these components is N:

$$\max_{i,k} \left| n_i^{(k)} \right| = N \tag{6}$$

Suppose that we have found a solution $\mathbf{p}^{\text{joint}}$ to equation (4) where $\mathbf{p}_{i,j}^{\text{joint}} > 0$ for all (i, j), and assume that

$$\min_{i,j} \left(\mathbf{p}_{i,j}^{\text{joint}} \right) = \varepsilon \tag{7}$$

and hence that

$$\max_{i,j} \left(\mathbf{p}_{i,j}^{\text{joint}} \right) = 1 - \varepsilon \tag{8}$$

If $\mathbf{p}^{\text{joint}}$ is a vector of joint probabilities that is a solution to equation (4), then any vector of the form

$$\hat{\mathbf{p}}^{\text{joint}}(a_1, a_2, a_3, a_4) = \mathbf{p}^{\text{joint}} + \sum_k a_k \mathbf{n}^{(k)} \tag{9}$$

will also be a solution to equation (4). In fact, any of the infinitely many vectors $\hat{\mathbf{p}}^{\text{joint}}(a_1, a_2, a_3, a_4)$ for which $a_k < \varepsilon / (4N)$ $(k = 1, 2, 3, 4)$ will be a solution, also satisfying

$$0 \le \hat{\mathbf{p}}^{\text{joint}} \le 1 \tag{10}$$

for all (i, j).

Above, we assume that we have a solution $\mathbf{p}^{\text{joint}}$ to equation (4) where $\mathbf{p}_{i,j}^{\text{joint}} > 0$ for all (i, j). However, the demonstration can easily be extended to the case where some of the components of $\mathbf{p}^{\text{joint}}$ are zero.

The conclusion is that further information (i.e. constraints) must be added in order to uniquely determine the joint probability from the marginal distributions.

Following a reasoning similar to the above example, it can easily be verified that if the inequality

$$LK^M + 1 < K^N \tag{11}$$

is satisfied, then the problem of determining a joint probability distribution from its marginal probability distributions (having the same dimension) is underdetermined (i.e. have some degree of freedom), and infinitely many joint distributions are possible. To ensure a unique solution to the problem, at least $K^N - LK^M - 1$ additional constraints must be supplied.

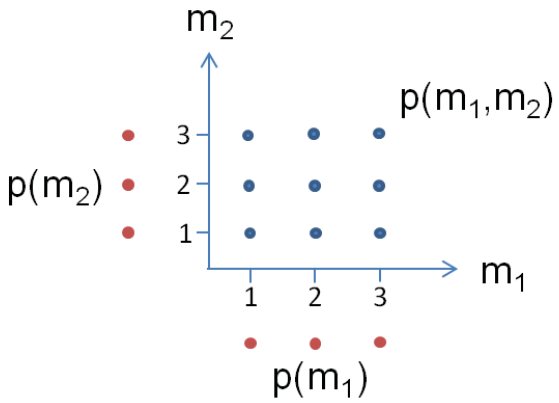


Figure 1. Outline of a two-dimensional joint probability distribution and two one-dimensional marginal probability distributions obtained from the joint probability distribution. The possible outcomes from the joint probability distribution are marked by blue dots and the outcomes from the marginal probability distributions are marked by red dots. In this particular example the number of possible outcomes from the marginal probability distribution is smaller than the number of possible outcomes from the joint probability distribution. It can be shown that infinitely many joint probability distributions are consistent with the given marginal distributions.

3 Adding constraints to obtain unique joint distributions

The degrees of freedom can be reduced if additional constrains can be imposed on the problem of determining a joint probability distribution based on its marginal probabilities. In what follows we will list some of the (not necessarily mutually exclusive) possibilities for reducing the degrees of freedom in the problem.

3.1 Assuming that the joint distribution is parametric

If the joint probability distribution is assumed to be parametric of some form, the degrees of freedom are reduced to the number of parameters that control the probability distribution. For instance, by assuming the joint probability distribution to be Gaussian reduces the degrees of freedom to $\frac{N^2+3N}{2}$ (N means, N variances, and $\frac{N(N-1)}{2}$ covariances).

3.2 Assuming that the joint distribution has Markov properties

Assuming that the joint probability distribution is Markovian also implies a reduction in the degrees of freedom. A Markovian probability distribution (or Markov random field) is defined as a random field in which probability distributions over the individual variables only depend on variable values within a certain neighborhood. Hence, a Markov random field is completely defined by a set of conditional probability distributions (one for each variable and its neighborhood) or, equivalently, by the corresponding set of marginal probability distributions. Hence, the number of degrees of freedom is equal the number of marginal probability distributions.

Assuming that the variables of the distribution are stochastically independent is a special case of the Markovian property in which the neighborhood of dependence of the individual variables is equal to the variables itself.

3.3 Assuming that the joint distribution is stationary

That a random field is stationary means that its neighborhoods do not change when shifted in space and their corresponding marginals are all equal to the same marginal probability distribution. In this way, when assuming the joint distribution to be Markovian, only a single marginal probability distribution is necessary to uniquely determine the joint probability distribution.

If the random field is assumed to be both Gaussian and stationary, the assumption of stationarity further reduces the degrees of freedom because in this case we only have one mean, one variance, and N covariances. Hence, the total degrees of freedom are reduced to $N+2$. In this case a semivariogram analysis of the sample model provides the variance and the covariances of the unknown joint probability distribution, and the semivariogram together with the mean of the sample model provides exactly the necessary information (N+2 degrees of freedom) in order to uniquely determine the unknown joint distribution.

Stationarity can also be exploited in cases where a sample model (often in form of a training image) from the joint probability distribution is available. A sample model is a set of

model parameters that is assumed to be a realization from the unknown joint probability distribution. By assuming that the underlying joint probability distribution represents a stationary process, all marginal probability distributions defined over model parameters related to a certain pixel configuration are equal. Under this assumption, the sample model provides a sample (i.e. multiple realizations) from the marginal probability distributions (related to a certain pixel configuration) of the unknown, underlying joint probability distribution.

By scanning the image with a template T_k (defined by a certain pixel configuration) an M-dimensional histogram with M-point statistical information from the sample model $\mathbf{m}^{SM}$ is provided. The pattern frequency distribution related to a template T, $\boldsymbol{\pi}(\mathbf{m}^{SM}, T)$, (i.e. normalized histogram) provides an estimate of an M-dimensional marginal probability distribution from the unknown joint probability distribution. The more times *L* that the template can be repeated across the image related to the sample model, the better becomes the approximation of the marginal probability distribution. See e.g. Strebelle (2002) for more details on how to obtain statistics from a training image.

3.5 Assuming that the joint distribution has maximum entropy

Assuming maximum entropy of a probability distribution is often considered the most "neutral" assumption possible. This assumption involves that among all possible joint probability distributions satisfying the marginal probability distributions, this probability distribution is the least biased estimate possible on the given information, i.e. it is maximally noncommittal with regard to missing information (Jaynes, 1957).

The entropy H of a variable $\mathbf{m}$ with probability distribution $p(\mathbf{m})$ is denoted as (Shannon, 1948)

$$H(\mathbf{m}) = \sum_i p(\mathbf{m}_i) \log_2 \left(p(\mathbf{m}_i) \right), \qquad (12)$$

where $\log_2$ is the logarithm with base 2. In this case the entropy is measured in bits. If nothing is known and assumed about the distribution, a uniform distribution is the one that carries least information and, therefore, has maximum entropy. Given the mean and covariance, the Gaussian distribution is the maximum entropy distribution. Hence, by assuming maximum entropy the degrees of freedom always match the information at hand and no degrees of freedom will be left.

3.5.1 Sequential simulation and entropy

Using the product rule, a joint probability distribution can be expressed as a product of conditional probability distributions

$$\begin{aligned} f(m_N,...,m_1) &= f(m_N \mid m_{N-1},...,m_1) f(m_{N-1},...,m_1) \\ &= f(m_N \mid m_{N-1},...,m_1) f(m_{N-1} \mid m_{N-2},...,m_1) f(m_{N-2},...,m_1) \\ &\vdots \\ &= \prod_{k=3}^{N} f(m_k \mid m_{k-1},...,m_1) f(m_2 \mid m_1) f(m_1) \end{aligned}, \qquad (13)$$

This equation is the backbone of sequential simulation, because in this way a realization of the joint distribution can be obtained by sequentially simulating the model parameters conditionally to all previously simulated model parameters (see e.g. Gómez-Hernández and Cassiraga (2000) and Hansen et al. (2012)).

The entropy of equation 13 is given as (Cover and Thomas, 2006)

$$H(m_N,...,m_1) = H(m_1) + H(m_2 \mid m_1) + \sum_{k=3}^{N} H(m_k \mid m_{k-1},...,m_1). \qquad (14)$$

It can be shown (Cover and Thomas, 2006) that conditioning leads to a decrease in entropy:

$$H(m_1 \mid m_2) \le H(m_1) \qquad (15)$$

with equality if m_1 and m_2 are independent. Consider now the case where not all marginal probability distributions can be provided (as is otherwise demanded in the formulation in equation 13), but only some of them can be obtained, which will be the case when the marginals are extracted from a training image (e.g. $p(m_5 \mid m_1, m_2, m_3, m_4)$ is required in the sequential simulation, but only $p(m_5 \mid m_1, m_2)$ is extracted). In such a case the entropy of this term in equation 14 will increase, according to the inequality in equation 15 (e.g. $H(m_5 \mid m_1, m_2, m_3, m_4, m_5) \le H(m_5 \mid m_1, m_2)$), which will make the resulting entropy of joint distribution in equation 13 increase. Hence, every time the dimension of a marginal probability distribution used for the sequential simulation is pruned (i.e. some variables are not used), the information content of the resulting joint probability distribution is reduced.

4 Effects from additional constraints

As we have seen above, if we are only acquainted with a limited set of approximate marginal probability distributions based on a sample model, we are facing an underdetermined problem of finding the joint probability distribution. In the following we discuss and analyze some unwanted consequences of adding additional constrains to this undetermined problem by using different assumptions about the joint probability distribution. We address these consequences and come up with alternative formulations. Finally, we formulate a solution that also takes into account the statistical uncertainty related to the given marginal probability distributions based on a sample model.

4.1 The use of pruned marginal distributions in sequential simulation

A common assumption used by the sequential simulation algorithms is that the values of the individual model parameters are only dependent on other model parameters within a certain neighborhood surrounding the location associated with the parameter. As discussed earlier, this assumption is known as the Markov property. If the joint distribution is also assumed to be stationary, then only a single marginal distribution must be known in order to uniquely determine this Markovian joint distribution.

As a sequential simulation algorithm proceeds, the model parameters are simulated from a local conditional probability distribution (based on the known marginal distribution) one at a time (e.g. Strebelle, 2002). As the simulation evolves, the algorithm forms a directed graph, where each simulated model parameter is associated with a node and the directed edges between the nodes represent probabilistic dependencies between the model parameters. Since each node is only visited once during the sequential simulation, no cycle will ever occur in such a graph. Hence, the graph is said to be a directed acyclic graph (DAG) (see e.g. Bishop (2006) for an introduction to graphical models).

Figure 2(I) shows a set of nodes marked by letters A – I to be sequentially simulated. Figure 2(II) shows a DAG that occurs as a result of sequential simulation with the simulation sequence marked by numbers on the nodes. The dependencies defined by the directed edges result from a square-shaped template of 3 by 3 interconnected nodes.

The joint probability distribution over a DAG can be factorized as follows (e.g. Whittaker, 1990)

$$f(m_1,\ldots,m_N)=\prod_{i=1}^{N}p(m_i\mid pa(m_i))=\alpha\prod_{i=1}^{N}p(m_i,pa(m_i)),\qquad(16)$$

with the normalization factor $\alpha=\left(\prod_{i=1}^{N}p(pa(m_i))\right)^{-1}$.

$pa(m_i)$ is the set of so-called parents of the model parameter m_i. As an example, the model parameters associated with the nodes D, E, G, and I in Fig 2(II) are parents of the model parameter associated with the node H. The product in equation 16 is evaluated in the same order as the model parameters are visited during the sequential simulation. From equation 16 we now have a general formulation of the joint probability distribution that results from a sequential simulation algorithm.

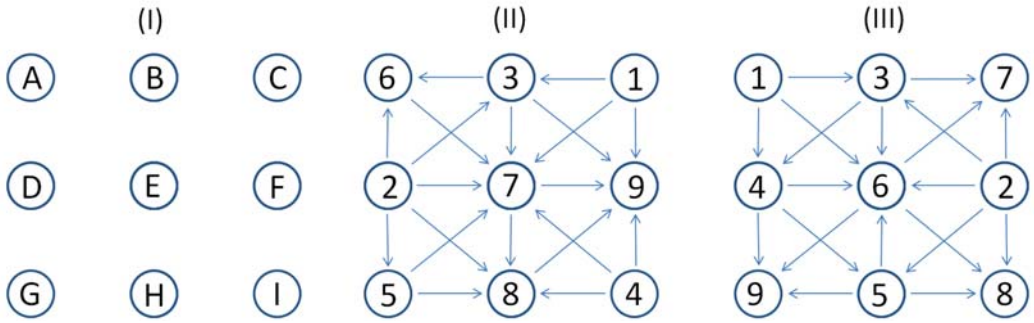


Figure 2. (I) Graph with 9 nodes labeled with letters A – I. (II) First example of sequential simulation sequence in a directed acyclic graph. The numbers are associated with the order by which the nodes are sequentially simulated. The arrows show the dependencies between the nodes as a result of a template of 3 by 3 interconnected nodes with the "central node" located in the center of the template. Hence, using this template the individual nodes can only be dependent on other nodes that are located next to them. (III) Second example of a sequential simulation path (as a result of the same template).

The joint probability distribution in equation 16 evaluated for the simulation sequences (seq. 1) as seen in Fig 2(II) is given as:

$$\begin{aligned}f_{seq1}(A,B,\ldots,I)&=p(C)p(D)p(B\mid C,D)p(I)p(G\mid D)p(A\mid B,D)\cdot\\&=p(E\mid A,B,C,D,G,I)p(H\mid D,E,G,I)p(F\mid B,C,E,H,I)\end{aligned}\qquad(17)$$

Using the product rule this expression can be rewritten as:

$$f_{seq1}(A,B,...,I) = \frac{p(C)p(B,C,D)p(I)p(G,D)p(A,B,D)p(E,A,B,C,D,G,I)}{p(C,D)p(B,D)p(A,B,C,D,G,I)p(D,E,G,I)p(B,C,E,H,I)} \cdot \\ = p(H,D,E,G,I)p(F,B,C,E,H,I) \quad (18)$$

Another choice of simulation sequence (seq. 2) is shown in Fig 2(III). The associated joint probability distribution over this graph is given as:

$$f_{seq2}(A,B,...,I) = p(A)p(F)p(B|A,F)p(D|A,B)p(H|D,F)p(E|A,B,D,F,H) \cdot \\ p(C|B,E,F)p(I|E,F,H)p(G|D,E,H) \\ = \frac{p(A)p(F)p(B,A,F)p(H,D,F)p(E,A,B,D,F,H)p(C,B,E,F)}{p(A,F)p(A,B)p(D,F)p(A,B,D,F,H)p(B,E,F)p(D,E,H)} \cdot \quad (19) \\ p(I,E,F,H)p(G,D,E,H)$$

From these examples (expressed in equations 18 and 19) it is seen that two different simulation sequences lead to two different joint probability distributions.

The above example shows that when Markov properties are assumed in a sequential simulation strategy as described above, the actual joint probability distribution sampled by the sequential simulation algorithm depends on (1): the size of the template T (typical for multiple-point statistic) or neighborhood (typical for two-point statistics), and (2): the simulation sequence. Note that if m_i always has all previously simulated nodes as its parents equation 16 equals equation 13 and is no longer dependent on the simulation sequence.

Equation 16 demonstrates that the joint probability distribution over a random simulation sequence is proportional to the product of marginal probability distributions over the individual nodes and their parents, $p\left(m_i, pa(m_i)\right)$. Since the set of model parameters $\left(m_i, pa(m_i)\right)$ are not necessarily related to the nodes associated with the configuration defined by template, this formulation of the joint probability distribution is not consistent with the known marginal probability distribution (i.e. does not satisfy equation 4).

4.1.1 Sequential simulation of the canonical joint distribution of a Markov random field

In an undirected graph a clique is defined as a subset of nodes in which every two nodes are connected by an edge. The marginal probability distributions from a joint distribution defined over such a graph contain the probabilistic dependencies over the cliques. The set of marginal probability distributions that is obtained from a sample model will be overlapping such that the

associated cliques will form a chain of cliques $(C_1,...,C_n)$. For a chain of cliques we have that: (1) every clique occurs exactly once and (2) $C_k \cap (C_1 \cup ... \cup C_{k-1})$ is a subset of at least one of the previous cliques $(C_1,...,C_{k-1})$. The set $S_k = C_k \cap (C_1 \cup ... \cup C_{k-1})$ is said to be the *k*-th separator. Hence, S_k is the overlap between clique C_k and the previous cliques in the chain. The set $R_k = C_k \setminus S_k$ is said to be the *k*-th residual.

The joint probability distribution defined over a chain of cliques is a Markov random field, which can be expressed as (see e.g. Castillo et al., 1997):

$$f(m_1,...,m_N) = \prod_{k=1}^{L} p(r_k \mid s_k) = \beta \prod_{k=1}^{L} p(r_k, s_k) = \beta \prod_{k=1}^{L} p(c_k), \tag{20}$$

where the normalization constant is given as $\beta = \left(\prod_{k=1}^{L} p(s_k) \right)^{-1}$. r_k and s_k are the model parameters associated with the node of the *k*'th set of residuals R_k and separators S_k, respectively. *L* is the total number of cliques in the chain of cliques. c_k are the model parameters associated with the nodes contained in a clique C_k, which is equal to the set of nodes given by the set (r_k, s_k).

In the graph depicted in Fig 3(I) an example of a chain of cliques is given by the subsets: $(ABDE, BCEF, DEGH, EFHI)$. The associated separators and residuals are given by the subsets $(\varnothing, BE, DE, EFH)$ and $(ABDE, CF, GH, I)$, respectively. The resulting joint probability distribution is then given as:

$$p(A,B,...,I) = \frac{p(A,B,D,E)\,p(B,C,E,F)\,p(D,E,G,H)\,p(E,F,H,I)}{p(B,E)\,p(D,E)\,p(E,F,H)}, \tag{21}$$

where the simulation sequence is seen in Fig 3(II). Note that the exemplified chain of cliques is not the only one possible.

Equation 20 is function of the known marginal distributions and is consistent with these marginals (i.e. satisfies equation 4). Therefore, this expression uses all the information provided by the known marginal probability distributions, which was not the case using the DAG formulation. Moreover, this expression guarantees that the same joint probability distribution (the canonical distribution) will be sampled from any simulation sequence that forms a chain of cliques from the set of cliques defined by the known marginal probability distributions. However, it should not be

forgotten that the nonuniqueness of the joint distribution consistent with given marginals still remains. The canonical distribution is just one particular choice out of infinitely many.

The formulation in equation 20 provides a means to perform sequential simulation from this Markovian joint distribution. As seen in Fig 3(II), sequential simulation based on this equation may at some points lead to sequential simulation of more than one model parameter at a time. However, this can easily be reformulated such that sequential simulation of one model parameter at a time can be obtained using the product rule (see equation 13) within each residual: For the *k*'th residual the conditional probability can be rewritten as:

$$p(r \mid s) = \frac{p(r,s)}{p(s)} = \frac{\prod_{j=3}^{J} p(r_j \mid r_{j-1},...,r_1,s)\, p(r_2 \mid r_1,s)\, p(r_1,s)}{p(s)} \tag{22}$$

where J is the number of nodes within the *k*'th set of residuals. This is an application of the product rule, which is not influenced by the simulation sequence.

Thus, a sequential simulation algorithm based on equations 20 and 22 will lead to an algorithm that generate realizations from a known joint probability distribution that is consistent with the known marginal probability distribution obtained from a sample model (such as a training image).

It should be noted that, for the Markovian joint distribution, the model with maximum probability will be the model only containing patterns with maximum probability in the marginal distribution due to the factorized formulation of the joint distribution (see equation 20). Hence, the maximum probability model may be very unrealistic and lack the variability and patterns as seen in the sample model.

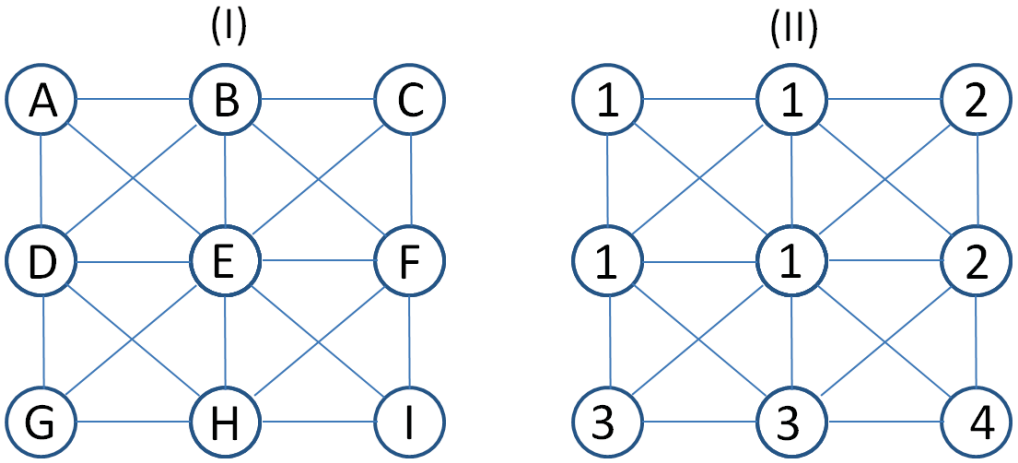


Figure 3. (I) An example of an undirected graph and the dependencies among the nodes marked by letter A, B,…, I. (II) The simulation sequence, when using a Markov random field formulation for the joint probability distribution over the nodes, is indicated by numbers.

4.2 Uncertain marginals

The sample model consists of a finite set of model parameters, which means that a marginal probability distribution $\boldsymbol{\pi}_{obs} = \boldsymbol{\pi}(\mathbf{m}^{SM}, T)$ estimated from such a sample is only an approximation to the true marginal distribution. We shall now consider how this statistical uncertainty can be taken into account in the Markovian joint distribution.

The probability that a certain pattern histogram $H = \boldsymbol{\pi} K^{\mathrm{Re}}$ is a realization from a process with outcome probabilities given by the observed pattern histogram from the sample model $H^{SM} = \pi_{obs} K^{SM}$ can be expressed by the Dirichlet distribution. Here, K^{Re} and K^{SM} are the number of counts in the pattern histograms related to a realization and the sample model (e.g. training image), respectively. If $\pi = (\pi_1, ..., \pi_K)$ is defined through $H = \boldsymbol{\pi} K^{\mathrm{Re}}$, this probability distribution is expressed as:

$$p(\pi_1, ..., \pi_K) = \frac{K^{\mathrm{Re}}!}{H_1!, ..., H_I!} \prod_{i=1}^{I} \left(\frac{H_i^{SM} + H_i^{prior}}{K^{SM} + K^{prior}} \right)^{H_i} \tag{23}$$

where $H_i = \pi_i K^{\mathrm{Re}}$, H_i^{SM} and H_i^{prior} are the *i*'th bins in the pattern histograms. H_i^{prior} is the prior histogram that expresses our prior expectations about the pattern histogram of the sample model. Typically, this histogram is assumed to be uniformly distributed because this is the information provided before the training image is seen. If multiple sample models are known, the resulting histogram is simply the sum of the pattern histograms related to the individual sample models. $I = K^M$ (i.e. K to the M'th power) is the number of possible pattern combinations within a certain template, which is a function of the template size M and the number of discrete outcomes from the categorical model parameters *K*.

Note that the Dirichlet distribution only needs to be evaluated for bins where $H_i \neq 0$. No other bins contribute to the probability. Hence, the histogram becomes sparse, which, in particular, saves memory for large template sizes and/or many categories of the model parameter values. In the limit $H_i^{prior} = 0, i = 1, 2, ..., I$ the Dirichlet distribution reduces to the multinomial distribution.

Each frequency distribution represents an approximation to a marginal probability distribution $p_{app} = \boldsymbol{\pi}$. Given an approximation to a marginal probability distribution we can define a Markovian joint probability distribution over a set of model parameters given a marginal probability distribution (using equation 20):

$$p(\mathbf{m} \mid \boldsymbol{\pi}) = \beta \prod_{k=1}^{L} p_{\mathrm{app}}(c_k) \tag{24}$$

From equation 23 and 24 we are then capable of defining a joint probability distribution over the model parameters and the frequency distributions as:

$$p(\mathbf{m}, \boldsymbol{\pi}) = p(\mathbf{m} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) \tag{25}$$

A marginalization over the frequency distribution gives us an expression for a Markovian joint probability distribution that takes into account the uncertainty related to the marginal probability distribution:

$$p(\mathbf{m}) = \sum_{\boldsymbol{\pi}} p(\mathbf{m}, \boldsymbol{\pi}) = \sum_{\boldsymbol{\pi}} p(\mathbf{m} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}) \tag{26}$$

Realizations from this distribution can be obtained by sequentially simulating from the joint probability distribution $p(\mathbf{m}, \boldsymbol{\pi})$ by:

1. Obtain a realization from the Dirichlet distribution $p(\boldsymbol{\pi})$.

2. Given a realization of $\boldsymbol{\pi}$, a realization of $\mathbf{m}$ can then be obtained through sequential simulation of $p(\mathbf{m}|\boldsymbol{\pi})$.

All realizations of $\mathbf{m}$ obtained in this way will then be realizations from $p(\mathbf{m})$.

Equation 26 is a formulation of the Markovian joint probability distribution, as defined in equation 20, that in addition takes into account the statistical uncertainty related to the approximate marginal probability distribution obtained from a sample model (e.g. due to the finite number of count that can be obtained from a training image). Hence, in such a formulation the exactness of the marginal probability, which is controlled by for example the size of the training image, has an influence on the variability of the joint probability distribution, which is not the case in existing approaches.

Taking this uncertainty into account is important in order to obtain a joint probability distribution that is not realistic, and over optimistic, with respect to the information content obtained from the sample model. In this way such a formulation can be used to provide a prior probability distribution that carry realistic prior information obtain from a training image rather than a "narrow" prior information that consider the training image to represent the exhaustive prior information about the subsurface.

4.3 Loss of information by algorithms with pruned marginal distributions

As we saw above, the joint probability distribution related to a DAG is not consistent with the known marginal distributions because $p(m_i, pa(m_i))$ was typically of a dimension smaller than the known marginal. On the other hand, for the Markovian joint probability distribution the probabilities over the cliques $p(r_k, s_k)$ are exactly the known marginal distribution. As we saw in section 4.5, when dependencies does not exist (i.e. are unknown or discarded), as is the case of the DAG based algorithms when compared to the canonical formulation in equation 20, the entropy of the joint distribution increases.

The Markovian joint probability distribution is consistent with the known marginal probability distributions. However, if only a set of lower-dimensional pruned marginal distributions is used, as is the case when using a DAG based algorithm (such as the SNESIM algorithm or SGSIM using a local neighborhood), the entropy of the resulting joint distribution that

is sampled by such algorithms is increased compared to the entropy of the Markovian joint distribution. In other words, information is lost when neighborhoods (i.e. marginal probability distribution) are pruned.

The SNESIM algorithm reduces the number of conditioning events if the denominator of the conditional probability distribution becomes zero (Strebelle, 2002). Pruning of a conditioning event leads to a joint distribution that is inconsistent with the marginals because a parent is removed (see equation 16), and leads to an increase in the entropy. Only if the simulation sequences and the pruning strategy are the same for two different SNESIM simulations, the joint distributions sampled during the two simulations will be the same.

When uncertainty of the marginal distributions is taken into account, as obtained through the Dirichlet distribution, the variability of the Markovian joint probability distribution will increase as compared to the case of not accounting for this uncertainty. This will, in turn, result in an increase in the entropy due to the less certain information provided by an uncertain marginal distribution.

5 Numerical examples

In this section we demonstrate different numerical examples. First, we calculate the entropy of the two joint probability distributions that are obtained using two different simulation sequences. Second, we see how this influences the probability of drawing a certain realization from these two joint probability distributions. Finally, we look at the relative error of using different sizes of neighborhood (ignoring different number of conditioning events) when performing sequential Gaussian simulation (SGSIM).

5.1 Quantifying the loss of information using different simulation sequences.

Figure 4A shows an example of a training image. Figure 4F shows the non-zero multiple-point-based probabilities extracted from the training image using the template seen in Fig 4B. Figures 4D and 4E are two examples of simulation sequences that can be used to obtain the realization of the model parameters $(a, b, c, d, e, f, g, h, i)$ seen in Fig 4C. The joint probability distribution resulting from simulation sequence no. 1 using equation 16 and the template with the central node as seen in Fig 4B is given as:

$$p_{seq1}(a,b,c,d,e,f,g,h,i) = p(e)p(b)p(i\mid e)p(a)p(h\mid e)p(f\mid b,e)p(g)p(c\mid b)p(d\mid a)$$
$$= p(e)p(b)\frac{p(i,e)}{p(e)}p(a)\frac{p(h,e)}{p(e)}\frac{p(f,b,e)}{p(b,e)}p(g)\frac{p(c,b)}{p(b)}\frac{p(d,a)}{p(a)}$$
$$= p(i,e)p(h,e)\frac{p(f,b,e)}{p(b,e)}p(c,b)p(d,a)$$

Whereas the joint distribution obtained when using simulation sequence no. 2 (for the same template) is given as:

$$p_{seq2}(a,b,c,d,e,f,h,h,i) = p(g)p(d)p(h\mid g,d)p(c)p(f\mid c)p(a)p(i\mid f,h)p(e\mid a,d)p(b\mid a)$$
$$= p(g)p(d)\frac{p(h,g,d)}{p(g,d)}p(c)\frac{p(f,c)}{p(c)}p(a)\frac{p(i,f,h)}{p(f,h)}\frac{p(e,a,d)}{p(a,d)}\frac{p(b,a)}{p(a)}$$
$$= p(g)p(d)\frac{p(h,g,d)}{p(g,d)}p(f,c)\frac{p(i,f,h)}{p(f,h)}\frac{p(e,a,d)}{p(a,d)}p(b,a)$$

The probability of obtaining the realization in Fig 4C using these two simulation sequences are

$$p_{seq1}(a=0,b=0,c=0,d=0,e=1,f=1,g=0,h=0,i=1) = \frac{5}{48}\frac{8}{48}\frac{4}{9}\frac{26}{48}\frac{24}{48} = 20.90\cdot 10^{-4}$$

and

$$p_{seq2}(a=0,b=0,c=0,d=0,e=1,f=1,g=0,h=0,i=1) = \frac{18}{48}\frac{18}{48}\frac{21}{24}\frac{6}{48}\frac{3}{5}\frac{4}{24}\frac{26}{48} = 8.33\cdot 10^{-4},$$

respectively.

The entropy of the variables (a,b,c,d,e,f,g,h,i) for the two different joint distributions is given as

$$
\begin{aligned}
H_{seq1}(a,b,c,d,e,f,g,h,i) &= H(e)+H(b)+H(i\mid e)+H(a)+H(h\mid e)+H(f\mid b,e)+H(g)+H(c\mid b)+H(d\mid a)\\
&= -\sum_{x=0}^{1} p(e=x)\log_2\, p(e=x) - \sum_{x=0}^{1} p(b=x)\log_2\, p(b=x) - \sum_{x=0}^{1} p(i=x\mid e)\log_2\, p(i=x\mid e)\\
&\quad - \sum_{x=0}^{1} p(a=x)\log_2\, p(a=x) - \sum_{x=0}^{1} p(h=x\mid e)\log_2\, p(h=x\mid e)\\
&\quad - \sum_{x=0}^{1} p(f=x\mid b,e)\log_2\, p(f=x\mid b,e) - \sum_{x=0}^{1} p(g=x)\log_2\, p(g=x)\\
&\quad - \sum_{x=0}^{1} p(c=x\mid b)\log_2\, p(c=x\mid b) - \sum_{x=0}^{1} p(d=x\mid a)\log_2\, p(d=x\mid a)\\
&= -4\left[\frac{18}{48}\log_2\left(\frac{18}{48}\right)+\left(1-\frac{18}{48}\right)\log_2\left(1-\frac{18}{48}\right)\right] - \frac{5}{14}\log_2\left(\frac{5}{14}\right)-\left(1-\frac{5}{14}\right)\log_2\left(1-\frac{5}{14}\right)\\
&\quad - \frac{8}{15}\log_2\left(\frac{8}{15}\right)-\left(1-\frac{8}{15}\right)\log_2\left(1-\frac{8}{15}\right) - \frac{7}{9}\log_2\left(\frac{7}{9}\right)-\left(1-\frac{7}{9}\right)\log_2\left(1-\frac{7}{9}\right)\\
&\quad - \frac{24}{32}\log_2\left(\frac{24}{32}\right)-\left(1-\frac{24}{32}\right)\log_2\left(1-\frac{24}{32}\right) - \frac{24}{33}\log_2\left(\frac{24}{33}\right)-\left(1-\frac{24}{33}\right)\log_2\left(1-\frac{24}{33}\right)\\
&= 4.79 bits
\end{aligned}
$$

and

$$
\begin{aligned}
H_{seq2}(a,b,c,d,e,f,g,h,i) &= H(g)+H(d)+H(h\mid g,d)+H(c)+H(f\mid c)+H(a)+H(i\mid f,h)+H(e\mid a,d)+H(b\mid a)\\
&= -\sum_{x=0}^{1} p(g=x)\log_2\, p(g=x) - \sum_{x=0}^{1} p(d=x)\log_2\, p(d=x) - \sum_{x=0}^{1} p(h=x\mid g,d)\log_2\, p(h=x\mid g,d)\\
&\quad - \sum_{x=0}^{1} p(c=x)\log_2\, p(c=x) - \sum_{x=0}^{1} p(f=x\mid c)\log_2\, p(f=x\mid c)\\
&\quad - \sum_{x=0}^{1} p(a=x)\log_2\, p(a=x) - \sum_{x=0}^{1} p(i=x\mid f,h)\log_2\, p(i=x\mid f,h)\\
&\quad - \sum_{x=0}^{1} p(e=x\mid a,d)\log_2\, p(e=x\mid a,d) - \sum_{x=0}^{1} p(b=x\mid a)\log_2\, p(b=x\mid a)\\
&= -4\left[\frac{18}{48}\log_2\left(\frac{18}{48}\right)+\left(1-\frac{18}{48}\right)\log_2\left(1-\frac{18}{48}\right)\right] - \frac{21}{25}\log_2\left(\frac{21}{25}\right)-\left(1-\frac{21}{25}\right)\log_2\left(1-\frac{21}{25}\right)\\
&\quad - \frac{9}{33}\log_2\left(\frac{9}{33}\right)-\left(1-\frac{9}{33}\right)\log_2\left(1-\frac{9}{33}\right) - \frac{3}{7}\log_2\left(\frac{3}{7}\right)-\left(1-\frac{3}{7}\right)\log_2\left(1-\frac{3}{7}\right)\\
&\quad - \frac{4}{25}\log_2\left(\frac{4}{25}\right)-\left(1-\frac{4}{25}\right)\log_2\left(1-\frac{4}{25}\right) - \frac{24}{30}\log_2\left(\frac{24}{30}\right)-\left(1-\frac{24}{30}\right)\log_2\left(1-\frac{24}{30}\right)\\
&= 4.25 bits
\end{aligned}
$$

, respectively.

Sequential simulation sequence no. 2 leads to a joint distribution that is based on conditional probability distributions that are conditional to more events than is the case for simulation sequence no. 1. Consequently, the entropy of the joint distribution related to sequence 1 is higher than for sequence no. 2. This means that more information will be propagated from the input training image to the output realizations using simulation sequence no. 2. The loss of information by choosing simulation sequence no. 1 instead for sequence no. 2 is $0.54 bits$. Hence,

the joint probability distribution based on simulation sequence no. 1 is more “flat” than the distribution based in simulation sequence no. 2. This is also reflected in the probabilities of the realization evaluated in the two different joint distributions, where the probability is lowest for the joint distributions with the highest entropy.

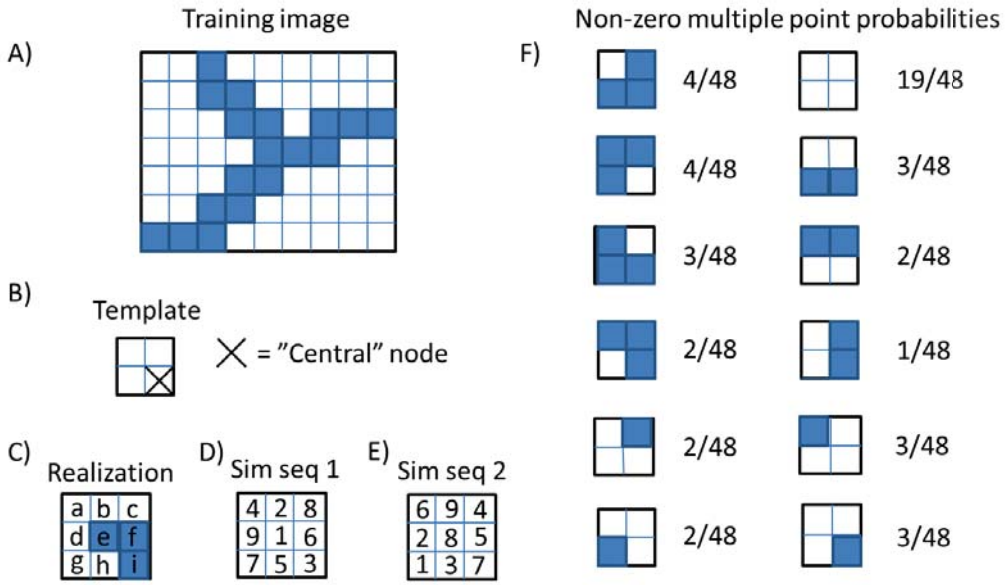


Figure 4. A) Training image. The white pixels refer to model parameters with value 0 and the blue pixels to the value 1. B) Template used to extract multiple-point statistics. C) An example of a realization. D) Sequential simulation sequence no. 1. E) Sequential simulation sequence no. 2. D) The non-zero multiple-point probabilities extracted from the training image using the displayed template.

5.2 Quantifying the relative error of using different neighborhoods

The relative error caused by using a (reference) joint probability distribution based on certain assumptions p_{ref} as compared to another formulation of the joint distribution p can be expressed as

$$E_{relative} = \frac{p_{ref}(\mathbf{m}) - p(\mathbf{m})}{p_{ref}(\mathbf{m})}, \qquad (27)$$

where $\mathbf{m}$ is a realization of the model parameters.

An example of using the above error measure is seen in Fig 5. Here, the probability of a realization from an exact Gaussian joint probability obtained through SGSIM (with no neighborhood) is compared with the probability of the same realization evaluated for the approximated formulation of the Gaussian distribution using a neighborhood (this amounts to the case of using only some marginals in equation 13 instead of using all marginals). In this way the error made when using the approximated formulation is quantified. The calculations are performed in a grid of 10 x 10 grid cell (i.e. 100 pixels) using an isotropic exponential covariance function with a range of 5 cells.

Figure 5 shows the relative error of using a local neighborhood that maximum includes the 50, 20 and 5 nearest conditioning model parameters (i.e. maximum number of parents in equation 16) relative to the probability of using no neighborhood (i.e. conditioning to all previously simulated model parameters). The result demonstrates that, in general, the larger the neighborhood the more accurate becomes the joint distribution (compare the individual curves). Moreover, the joint distribution is seen to depend on the simulation sequence (i.e. random path used), which is seen as the variability of the individual curves.

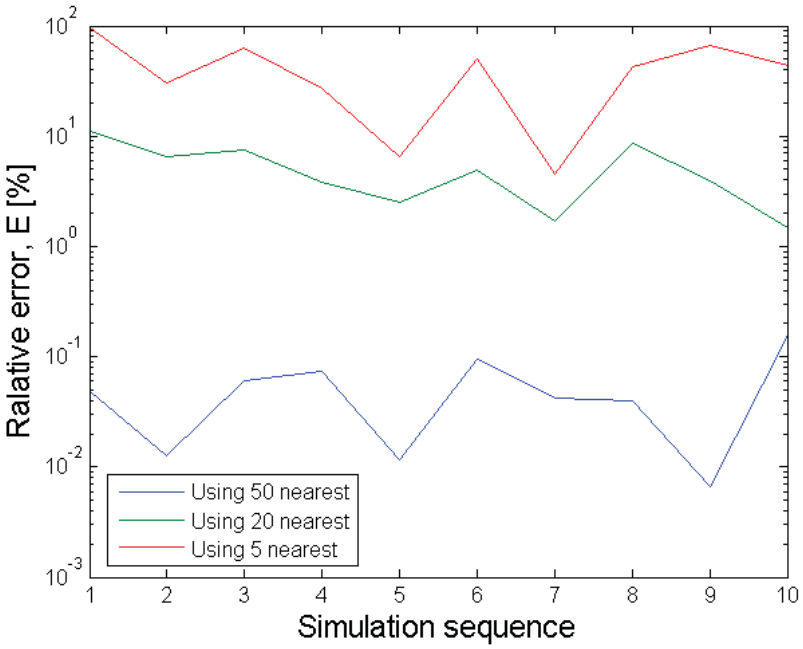


Figure 5. The relative error of realizations obtained from approximate Gaussian probability distributions using different neighborhoods that include the 50, 20 and 5 nearest conditioning events, respectively, as compared to the case of using all conditioning events. The error is calculated for 10 different random simulation sequences (i.e. random seeds). It is seen that the relative error is influences by the number of conditioning events used (the three different curves) and by the simulation sequence used (the variability of the individual curves).

6 Conclusion

The problem of determining a joint probability distribution based on a limited set of its marginal probability distributions is in general an underdetermined problem. Different assumptions about the joint probability distribution can be made that reduce the degrees of freedom of this problem. The degrees of freedom of a general formulation of such a problem have been quantified and the influence of these assumptions has been discussed.

In practice a set of marginal probability distribution from an unknown joint probability distribution can be estimated based on a realization (i.e. referred to as a sample model) from the joint probability distribution under the assumption of stationarity. We have investigated the

consequences of using different assumptions in order to uniquely determine a joint probability from its marginal probability distributions based on a single realization.

The implicit assumptions made by typical sequential simulation algorithms have been investigated and the joint probability distribution that is sampled by such algorithms has been expressed. We find that these joint probability distributions depend on the simulation sequence (i.e. the random path used) and the size of the template/neighborhood. Additionally, this joint probability distribution is generally not consistent with the known marginal probability distributions. Moreover, the information content, measured as the entropy in the resulting joint probability distributions, decreases when the known marginal probability distributions are pruned during the sequential simulation.

From Markov random field theory a solution that provides a means of defining a Markovian joint probability distribution that is consistent with the known marginal probability distribution is provided. At the same time this formulation can also be used for sequential simulation that leads to realizations from the same joint probability distribution independent of the simulation sequence. The simulation sequence is not related to the individual nodes, but to a chain of cliques overlapping at least one of the previous cliques in the sequence.

Finally, we formulate a sequential simulation algorithm that takes into account the uncertainty related to marginal probability distributions estimated from sample models.

Acknowledgements

We wish to thank DONG E&P for financial support.

References

Bishop, CM (2006) Pattern recognition and machine learning. Springer Science, New York.

Castillo E, Gutiérrez JM, Hadi AS (1997) Expert systems and probabilistic network models. Springer-Verlag New York, Inc.

Cordua KS, Hansen TM, Mosegaard K (2012) Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information. Geophysics, 77: H19 – H31.

Cover T, Thomas J (2006) Elements of Information Theory. John Wiley & Sons, Inc., Hoboken, New Jersey.

Deutsch CV, Wang L (1996) Hierarchical object-based geostatistical modeling of fluvial reservoirs. Paper SPE 36514.

Gómez-Hernández J, Cassiraga EF (2000) Sequential conditional simulations with linear constrains. In: Monestiez P, Allard D, Froideveaux R (Eds.), Geostatistics 2000 Cape Town. Geostatistical Association of Southern Africa.

Guardiano F, Srivastava M (1993) Multivariate geostatistics: beyond bivariate moments. In: Geostatistics - Troia, pp. 133 – 144. Kluwer Academic, Dordrecht.

Hansen TM, Cordua KS, Mosegaard K (2012) Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. Computational Geosciences, 16: 593 – 611. doi:10.1007/s/10596-011-9271-1.

Jaynes ET (1957) Information theory and statistical mechanics. The Physical Review, 106(4): 620 - 630.

Journel AG (1994) Modeling uncertainty: some conceptual thoughts. *in* Dimitrakopoulos, R., ed., Geostatistics for the Next Century: Kluwer Academic Pub., Dordrecht, The Netherlands, p. 30–43.

Journel AG, Alabert F (1989) Non-Gaussian data expansion in the earth science. Terra Nova, v. 1:123 – 134.

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press.

Journel AG, Zhang T (2006) The necessity of a multiple-point prior model. Mathematical Geology, 38(5): 591–610.doi:10.1007/s11004-006-9031-2.

Lange K, Frydendall J, Cordua KS, Hansen TM, Melnikova Y, Mosegaard K (2012) A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information. Mathematical geosciences, 44: 783 – 803.

Looms MC, Hansen TM, Cordua KS, Nielsen L, Jensen KH, Binley A (2010) Geostatistical inference using crosshole ground-penetrating radar. Geophysics, 75(6): J29 – J41.

Mosegaard K, Singh S, Snyder D, Wagner H (1997) Monte Carlo analysis of seismic reflections from Moho and the W reflector. Journal of geophysical research, 102(B2): 2969 - 2981.

Shannon CE (1948) A Mathematical Theory of Communication. The Bell System Technical Journal. 27: 379 – 423.

Stien M, Kolbjørnsen O (2011) Facies Modeling Using a Markov Mesh Model Specification. Mathematical Geosciences, 43: 611 – 624.

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 Applied Mathematics.

Tjelmeland H and Besag J (1998) Markov random fields with higher order interactions. Scandanavian Journal of Statistics, 25: 415 – 433.

Whittaker J (1990) Graphical Models in Applied Multivariate Statistics. Wiley.

Appendix A8:
A Frequency Matching Method for Generation of a Priori Sample Models from Training Images

Authors:

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

Published in:

Proceedings of the 15th Annual Conference of the International Association for Mathematical Geoscience (IAMG 2011)

Salzburg, Austria

5 – 9 September 2011

A Frequency Matching Method for Generation of a Priori Sample Models from Training Images

Katrine LANGE[1], Knud Skou CORDUA[1], Jan FRYDENDALL[1], Thomas Mejer HANSEN[1], and Klaus MOSEGAARD[1]

[1] Center of Energy Resources Engineering,
Department of Informatics and Mathematical Modeling,
Technical University of Denmark, Denmark,
katla@imm.dtu.dk

Peer-reviewed IAMG 2011 publication

doi:10.5242/iamg.2011.0262

Abstract

This paper presents a Frequency Matching Method (FMM) for generation of a priori sample models based on training images and illustrates its use by an example. In geostatistics, training images are used to represent a priori knowledge or expectations of models, and the FMM can be used to generate new images that share the same multi-point statistics as a given training image.

The FMM proceeds by iteratively updating voxel values of an image until the frequency of patterns in the image matches the frequency of patterns in the training image; making the resulting image statistically indistinguishable from the training image.

1. Background

Consider a training image with N voxels (or pixels if the image is only 2D). Let z_k denote the value of the kth voxel of the image, $k = 1, \dots, N$. Here, we shall assume that the training image is a realization of a random process satisfying:

1) Voxel value z_k depends only on the values of the voxels in a certain neighborhood $\mathcal{N}_k$ around voxel k. Voxel k itself is not contained in $\mathcal{N}_k$. Let $\mathbf{z}_k$ be an ordered vector of the values of the voxels in $\mathcal{N}_k$; we then have:

$$f_Z\left(z_k \middle| z_N, \dots, z_{k+1}, z_{k-1}, \dots, z_1\right) = f_Z(z_k|\mathbf{z}_k).$$

2) For an image of infinite size the geometrical shape of all neighborhoods $\mathcal{N}_k$ are identical. This implies that if voxel k has coordinates (k_1, k_2, k_3), and voxel l has coordinates (l_1, l_2, l_3), then:

$$(n_1, n_2, n_3) \in \mathcal{N}_k \Rightarrow (n_1 - k_1 + l_1, n_2 - k_2 + l_2, n_3 - k_3 + l_3) \in \mathcal{N}_l.$$

3) We assume ergodicity, i.e.:

$$\mathbf{z}_k = \mathbf{z}_l \Rightarrow f_Z(z_k|\mathbf{z}_k) = f_Z(z_l|\mathbf{z}_l).$$

The basis of sequential simulation (e.g. Strebelle, 2002) is to exploit the assumptions above to estimate $f_Z(z_k|\mathbf{z}_k)$, and to use these conditions to generate new realizations of the random process from which the training image is a realization. The FMM does not operate by directly using conditional probabilities but it represents images by their frequency distribution, which is derived using neighborhoods of voxels. The frequency distribution is closely related to conditional probabilities.

2. The Frequency Distribution

Before presenting the FMM we need to define what we denote the frequency distribution. To do so we will reuse the concept of neighborhoods from section 1 as well as the notation. Given an image with the set of voxels $\mathcal{Z} = \{1, \dots, N\}$ and voxel values $z_1, \dots, z_N$ we define the template function Ω as a function that takes as argument a voxel k and returns the set of voxels belonging to the neighborhood of voxel k. The neighborhood is denoted $\mathcal{N}_k$, and we will use the notation $\mathcal{N}_k = \Omega(k)$.

In the FMM the neighborhood of a voxel is indirectly given by the statistical properties of the image itself; however, the shape of a neighborhood satisfying the assumptions from section 1 is unknown. For each training image one must therefore define a template function that seeks to correctly describe the neighborhood.

Let $|\mathcal{N}_k|$ denote the number of voxels in $\mathcal{N}_k$. We define the set of inner voxels, $\mathcal{Z}_{in}$, of the image as:

$$\mathcal{Z}_{in} = \left\{k \middle| |\mathcal{N}_k| = \max_{l \in \mathcal{Z}} |\mathcal{N}_l| \right\}.$$

Typically, voxels on the boundary or close to the boundary of an image will not be inner voxels. It is the choice of template function that determines whether or not a voxel is an inner voxel.

The frequency distribution of an image is computed by scanning through all inner voxels of the image. For each of these we identify first the neighboring voxels and then the values of those. For voxel $k \in \mathcal{Z}_{in}$, the values of the neighboring voxels are denoted by the vector $\mathbf{z}_k$. The length of this vector equals the number of voxels in the neighborhood $\mathcal{N}_k$, which will be constant for all inner voxels; this follows trivially from the definition of inner voxels. We denote this number n. As each voxel can take on m different values, there exists up to m^n different types of neighborhoods; i.e. m^n different combinations for the values in $\mathbf{z}_k$.

Using the above definition of a neighborhood we now introduce the concept of patterns. The kth pattern $\mathcal{P}_k$ of the image is defined as the union of an inner voxel k and the set of its neighboring voxels. We will denote voxel k the center voxel of the kth pattern regardless of the geometrical shape of $\mathcal{P}_k$. Trivially, it follows that there exist m^{n+1} different types of patterns in the image. The type of a pattern is characterized by the (ordered) values of $\mathbf{z}_k$ and the value of the kth voxel itself. It should be stressed that the subindex k of $\mathcal{P}_k$, as well as of $\mathcal{N}_k$, represents the center voxel and thereby the location of the pattern, and it does not contrain any information on the type of the pattern.

Let p_i, for $i = 1, \dots, m^{n+1}$, count the number of times a pattern of type i appears in the image. These counts are used to represent the frequency distribution of an image. After having scanned through all inner voxels exactly once (the order is irrelevant) the frequency distribution is given by the vector $\mathbf{p}$:

$$\mathbf{p} = [p_1 \quad \cdots \quad p_{m^{n+1}}] = p_\Omega(z_1, \ldots, z_N).$$

Here p_Ω is the function that, given an image and a template function Ω, computes the frequency distribution of the image with respect to the template as just described.

We notice that, for a given template, the frequency distribution of an image is uniquely determined. The opposite, however, does not hold. Different images can have the same frequency distribution. This is exactly what we seek to exploit by using the frequency distribution to generate multiple new images, at the same time similar to, and different from, our training image.

3 The Frequency Matching Method

The Frequency Matching Method proceeds by iteratively updating voxel values of an image, until the frequency of patterns in the image matches the frequency patterns in the training image. One of the primary tasks when formulating the method is to define a similarity function for how close the frequency distributions of two images are. Below we shall define the similarity function used in the current implementation, and describe the optimization method we have applied to solve the combinatorial optimization problem arising from this.

3.1 The Similarity Function

The similarity function plays the following two important roles:

I. It allows us to determine if the frequency distribution of an image and the frequency distribution of a training image are identical within the accuracy required and we therefore consider the image a valid realization of the random process from which the training image is a realized.

II. Given two different images, no matter how similar they might be, and a training image, the similarity function should determine which of the two images is most similar to a valid realization of the same process as the training image, or if the two images are equally similar. At the same time it should reflect (in some sense) how close the images are to being a valid realization.

Using an iterative solution method, point I is used to determine if the method has converged to an acceptable solution, whereas point II guides the method through the solution space, helping it to converge.

As we do not know the random process of which the training image is a realization, we have chosen the chi-square measure of 'goodness of fit' between two sets of nominal data as a similarity function for our FMM implementation. This measure determines the distance between to frequency distributions by comparing the proportions of types of pattern in the two.

3.2 Applying the χ^2 Measure in the FMM

The chi-square measure can be applied to our situation using the following interpretations (see Bere and Chimedza, 2007):

samples	Each frequency distribution is considered a sample, i.e., we have two independent samples; one for the image itself and one for the training image.
categories	The samples are categorized with respect to the m^{n+1} exclusive and exhaustive types of patterns.
observations	Each appearance or count of a pattern is an observation. For each sample, the number of observations equals the number of inner

voxels in the corresponding image.

Given the frequency distributions of an image, $\mathbf{p}$, and of a training image, $\boldsymbol{\pi}$, we can compute what we denote to be the similarity function value of the image:

$$f(\mathbf{p}) = \chi^2(\mathbf{p},\boldsymbol{\pi}) = \sum_{i=1}^{m^{n+1}} \frac{(p_i - e_i)^2}{e_i} + \sum_{i=1}^{m^{n+1}} \frac{(\pi_i - \varepsilon_i)^2}{\varepsilon_i},$$

where e_i and ε_i denote the expected count of patterns of type i of the image and the training image, respectively. These are computed as:

$$e_i = \frac{(p_i + \pi_i)}{n_p + n_\pi} n_p,$$

$$\epsilon_i = \frac{(p_i + \pi_i)}{n_p + n_\pi} n_\pi,$$

and n_p and n_π are the number of inner voxels in the image and the training image, respectively.

Let χ^2 denote the chi-square value of the image computed from the two frequency distributions $\mathbf{p}$ and $\boldsymbol{\pi}$. f is a function of the frequency distribution $\mathbf{p}$ of the image, and the frequency distribution $\boldsymbol{\pi}$ of the training image. The training image and therefore its frequency distribution will remain unchanged when computing a new image; $\boldsymbol{\pi}$ has therefore been omitted as an argument of the similarity function. Furthermore, the frequency distribution $\mathbf{p}$ of the image is derived given a template function, i.e., the argument $\mathbf{p}$ of f depends on a template as well as on $z_1, \dots, z_N$, which means f is in fact a function of the image and a template function. However, to simplify the text, we have chosen to avoid these dependencies in the notation.

3.3 The Optimization Problem

The function f defined in section 3.2 seems to fulfill the two requirements we had, making the FMM a combinatorial optimization problem. The variables are the voxel values of the image. They can take on m different integer values namely $\{0, \dots, m-1\}$. Binary images, for instance, have $m = 2$. Given a template function Ω, the frequency distributions of the solution image, $\mathbf{p}$, and of a training image, $\boldsymbol{\pi}$, are computed by the frequency function p_Ω. Based on the two frequency distributions the similarity function of the image is computed. By minimizing the similarity function with respect to certain constraints, we can create images sharing the same multiple-point statistics as the training image. The resulting optimization problem can be expressed as follows:

$$\begin{aligned} \min_{z_1,..,z_N} \quad & f(\mathbf{p}) \\ \text{w.r.t.} \quad & \mathbf{p} = p_\Omega(z_1, \dots, z_N), \\ & z_k \in \{0, \dots, m-1\}, \quad \text{for } k = 1, \dots, N. \end{aligned}$$

If some of the voxel values are known beforehand, and the voxels are therefore not free variables, the last set of constraints can easily be altered, such that the set of values that the $\boldsymbol{k}$th voxel can take is only a subset of $\{\mathbf{0}, \dots, \boldsymbol{m-1}\}$.

3. Example

We have now introduced the Frequency Matching Method for generating a priori sample models from training images, and this has led us to a combinatorial optimization problem. Our choice of solution method is, for now, the intuitively simple heuristic Simulated Annealing (SA) (e.g. Kirkpatrick et al., 1983). For future work we would also like to explore other solution methods in the hope of finding one better suited for optimization and sampling problems.

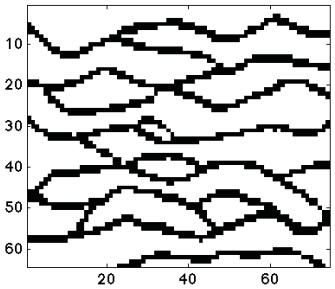


Figure 1: Training image.

The FMM has been implemented in MATLAB. To demonstrate the FMM we will consider a two-dimensional, binary training image with channel structures, see Figure 1.

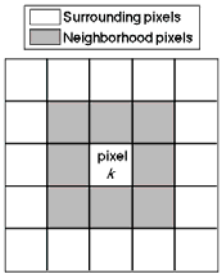


Figure 2: The template

We have defined the template such that the neighborhood of an arbitrary inner pixel k contains exactly the eight nearest pixels, see Figure 2. This relatively small neighborhood size is unlikely to completely satisfy our assumption of a pixel only being conditioned upon the pixels in its neighborhood. However, it will be shown that the method is still able to compute an acceptable solution. Due to the complexity of the method the size of the neighborhood greatly influences the running times, and for using much bigger templates we recommend implementing the method in Fortran, for instance.

We choose the exponential cooling rate for the SA, and the algorithm parameters are chosen manually. Discussing the strategies for choosing these optimally is beyond the scope of this text.

The starting image for SA is chosen to be all white. The SA algorithm searches the solution space consisting of images, and it moves from one image to another by randomly choosing a pixel and changing its value. Figure 3 and Figure 4 show the normalized frequency distributions of the training image and the image computed by the FMM, respectively. By 'normalized' we mean relative to the number of inner pixels in each of the images. Any normalized frequency distribution therefore sums to 1. Here we have truncated the ordinates of Figure 3 and Figure 4, as only one entry is significantly bigger than 0.08. The last entry is approximately 0.42 for both images. This entry is the one representing a white center pixel surrounded by all white neighboring pixels.

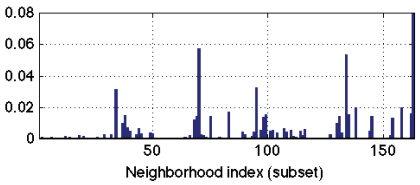


Figure 3: Normalized frequency distribution of the training image.

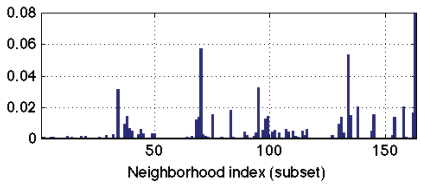


Figure 4: Normalized frequency distribution of the optimal solution image.

Notice that in Figure 3 and Figure 4 indexes corresponding to types of patterns appearing in neither the training image nor the solution image have been omitted

We observe that the FMM in terms of the frequency distributions has managed to match the training image quite well. Summing the bars of Figure 5 reveals that the two images have approximately 96.7% of their patterns in common. This number is likely to be improved by changing the parameters of the SA algorithm.

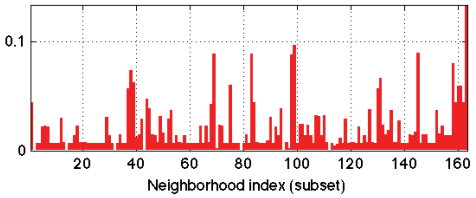


Figure 5: The absolute difference (in percent) between the normalized frequency distributions in Figure 3 and Figure 4.

Keep in mind that matching the frequency distributions only results in a useful image if our assumptions are met; i.e., if we chose a suitable template. Choosing too big a template means very long running times without sufficient gain in accuracy, and choosing too small a template will result in the picture not being similar to the training image. Our choice seems sufficient although not perfect, see Figure 6.

Figure 6 shows the image computed by the FMM. For this test case we have chosen to compute a 60×60 image based on a 64×74 training image but the method can produce images of arbitrary size. We notice that despite the relatively small template size, we have successfully recreated the channel structures. The channels even occasionally form loops, just like the channels of the training image.

One significant difference between the computed image and the training image is that the channels in the computed image are not all horizontally continuous across the image. We expect that this is merely a matter of choice of template and also the number of iterations the algorithm has been allowed to perform.

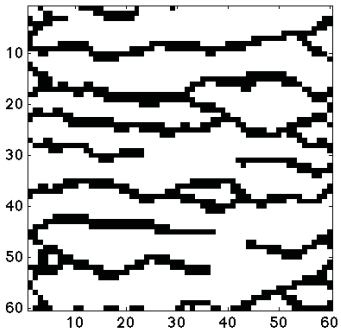


Figure 6: The computed solution image.

Another difference is the boundaries. It seems the method creates some artifacts along the boundaries. The density of channels is much higher on the left and right boundary then in the middle of the image. In the middle it resembles our training image and we therefore could have some issues in the way we treat non-inner pixels.

Notice how matching the frequency distributions indirectly results in the proportion of channels versus background in the computed picture to be in correspondence with the proportion of channels versus background in the training image. As stated, this is merely an example of the performance of the FMM. The method has also been applied to training images with different structures and shown similar results.

4. Conclusions and Future Work

In this paper we have derived the Frequency Matching Method for generation of a priori sample models from training images. We have implemented the method in MATLAB and shown the results of a simplified test case. The test example shows that the method is indeed able to produce an image that shares the same multi-point statistics as the training image.

This paper only scratches the surface of this newly developed method. In order to better understand its potential we would like to:

- Experiment thoroughly with training images with different structures.
- Investigate the convergence rate and performance of the FMM combined with other optimization methods.
- Explain and eventually avoid possible artifacts for non-inner voxels.

References

BERE, A., CHIMEDZA, C. (2007): A Comparative Study of the Accuracy of the Chi-Squared Approximation for the Power-Divergence Statistic and Pearson's Chi-Square Statistic in Sparse Contingency Tables. Journal of Statistical Research, Vol. 41, No. 2, pp. 73-81.

KIRKPATRICK, S., GELATT, C. D., VECCHI, M. P. (1983): Optimization by Simulated Annealing. Science, New Series, Vol. 220, No. 4598, pp. 671-680, May.

STREBELLE, S. (2002): Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. Mathematical Geology, Vol. 34, No. 1, January.

Appendix A9:

A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information

Authors:

Katrine Lange, Jan Frydendall, Knud Skou Cordua, Thomas Mejer Hansen, Yulia Melnikova, and Klaus Mosegaard

Published in:

Mathematical Geosciences

Math Geosci (2012) 44:783–803
DOI 10.1007/s11004-012-9417-2

A Frequency Matching Method: Solving Inverse Problems by Use of Geologically Realistic Prior Information

Katrine Lange · Jan Frydendall · Knud Skou Cordua · Thomas Mejer Hansen · Yulia Melnikova · Klaus Mosegaard

Received: 3 February 2012 / Accepted: 16 July 2012 / Published online: 22 September 2012

Abstract The frequency matching method defines a closed form expression for a complex prior that quantifies the higher order statistics of a proposed solution model to an inverse problem. While existing solution methods to inverse problems are capable of sampling the solution space while taking into account arbitrarily complex a priori information defined by sample algorithms, it is not possible to directly compute the maximum a posteriori model, as the prior probability of a solution model cannot be expressed. We demonstrate how the frequency matching method enables us to compute the maximum a posteriori solution model to an inverse problem by using a priori information based on multiple point statistics learned from training images. We demonstrate the applicability of the suggested method on a synthetic tomographic crosshole inverse problem.

Keywords Geostatistics · Multiple point statistics · Training image · Maximum a posteriori solution

1 Introduction

Inverse problems arising in the field of geoscience are typically ill-posed; the available data are scarce and the solution to the inverse problem is therefore not well-determined. In probabilistic inverse problem theory the solution to a problem is given as an a posteriori probability density function that combines states of information provided by observed data and the a priori information (Tarantola 2005). The ambiguities of the solution of the inverse problem due to the lack of restrictions on the solution is then reflected in the a posteriori probability.

K. Lange (✉) · J. Frydendall · K.S. Cordua · T.M. Hansen · Y. Melnikova · K. Mosegaard
Center for Energy Resources Engineering, Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Kongens Lyngby, Denmark
e-mail: katla@imm.dtu.dk

A priori information used in probabilistic inverse problem theory is often covariance-based a priori models. In these models the spatial correlation between the model parameters is defined by two-point statistics. In reality, two-point-based a priori models are too limited to capture curvilinear features such as channels or cross beddings. It is therefore often insufficient to rely only on the two-point statistics, and thus higher order statistics must also be taken into account in order to correctly produce geologically realistic descriptions of the subsurface. It is assumed that geological information is available in the form of a training image. This image could for instance have been artificially created to describe the expectations for the solution model or it could be information from a previous solution to a comparable inverse problem. The computed models should not be identical to the training image, but rather express a compromise between honoring observed data and comply with the information extracted from the training image. The latter can be achieved by ensuring that the models have the same multiple point statistics as the training image.

Guardiano and Srivastava (1993) proposed a sequential simulation algorithm that was capable of simulating spatial features inferred from a training image. Their approach was computationally infeasible until Strebelle (2002) developed the single normal equation simulation (*snesim*) algorithm. Multiple point statistics in general and the *snesim* algorithm in particular have been widely used for creating models based on training images and for solving inverse problems, see for instance Caers and Zhang (2004), Arpat (2005), Hansen et al. (2008), Peredo and Ortiz (2010), Suzuki and Caers (2008), Jafarpour and Khodabakhshi (2011). A method called the probability perturbation method (PPM) has been proposed by Caers and Hoffman (2006). It allows for gradual deformation of one realization of *snesim* to another realization of *snesim*. Caers and Hoffman propose to use the PPM method to find a solution to an inverse problem that is consistent with both a complex prior model, as defined by a training image, and data observations. PPM is used iteratively to perturb a realization from *snesim* while reducing the data misfit. However, as demonstrated by Hansen et al. (2012), as a result of the probability of the prior model not being evaluated, the model found using PPM is not the maximizer of the posterior density function, but simply the realization of the multiple point based prior with the highest likelihood value. There is no control of how reasonable the computed model is with respect to the prior model. It may be highly unrealistic.

The sequential Gibbs sampling method by Hansen et al. (2012) is used to sample the a posteriori probability density function given, for example a training image based prior. However, as with the PPM it cannot be used for optimization and locating the maximum a posteriori (MAP) model, as the prior probability is not quantified. The focus of our research is the development of the frequency matching (FM) method. The core of this method is the characterization of images by their multiple point statistics. An image is represented by the histogram of the multiple point-based spatial event in the image; this histogram is denoted the frequency distribution of the image. The most significant aspect of this method, compared to existing methods based on multiple point statistics for solving inverse problems, is the fact that it explicitly formulates an a priori probability density distribution, which enables it to efficiently quantify the probability of a realization from the a priori probability.

The classical approach when solving inverse problems by the least squares methods assumes a Gaussian prior distribution with a certain expectation. Solution models

to the inverse problem are penalized depending on their deviation from the expected model. In the FM method, the frequency distribution of the training image acts as the expected model and a solution image is penalized depending on how much its frequency distribution deviates from that of the training image. To perform this comparison we introduce a dissimilarity measure between a training image and a model image as the χ^2 distance between their frequency distributions. Using this dissimilarity measure for quantifying the a priori probability of a model the FM method allows us to directly compute the MAP model, which is not possible using known techniques such as the PPM and sequential Gibbs sampling methods.

Another class of methods are the Markov random fields (MRF) methods (Tjelmeland and Besag 1998). The prior probability density given by Markov methods involves a product of a large number of marginals. A disadvantage is therefore, despite having an expression for the normalization constant, that it can be computationally expensive to compute. Subclasses of the MRF methods such as Markov mesh models (Stien and Kolbjørnsen 2011) and partially ordered Markov models (Cressie and Davidson 1998) avoid the computation of the normalization constant, and this advantage over the MRF methods is shared by the FM method. Moreover, in contrast to methods such as PMM and MRF, the FM method is fully non-parametric, as it does not require probability distributions to be written in a closed form.

This paper is ordered as follows. In Sect. 2 we define how we characterize images by their frequency distributions, we introduce our choice of a priori distribution of the inverse problem and we elaborate on how it can be incorporated into traditional inverse problem theory. Our implementation of the FM method is discussed in Sect. 3. In Sect. 4 we present our test case and the results when solving an inverse problem using frequency matching-based a priori information. Section 5 summarizes our findings and conclusions.

2 Method

In geosciences, inverse problems involve a set of measurements or observations $\mathbf{d}^{\text{obs}}$ used to determine the spatial distribution of physical properties of the subsurface. These properties are typically described by a model with a discrete set of parameters, $\mathbf{m}$. For simplicity, we will assume that the physical property is modeled using a regular grid in space. The model parameters are said to form an image of the physical property.

Consider the general forward problem,

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

of computing the observations $\mathbf{d}$ given the perhaps non-linear forward operator g and the model parameters $\mathbf{m}$. The values of the observation parameters are computed straightforwardly by applying the forward operator to the model parameters. The associated inverse problem consists of computing the model parameters $\mathbf{m}$ given the forward operator g and a set of observations $\mathbf{d}^{\text{obs}}$. As the inverse problem is usually severely under-determined, the model $\mathbf{m}$ that satisfies $\mathbf{d}^{\text{obs}} = g(\mathbf{m})$ is not uniquely determined. Furthermore, some of the models satisfying $\mathbf{d}^{\text{obs}} = g(\mathbf{m})$ within the required level of accuracy will be uninteresting for a geoscientist as the nature of the

forward operator g and the measurement noise in $\mathbf{d}^{\text{obs}}$ may yield a physically unrealistic description of the property. The inverse problem therefore consists of not just computing a set of model parameters satisfying Eq. 1, but computing a set of model parameters that gives a realistic description of the physical property while honoring the observed data. The FM method is used to express how geologically reasonable a model is by quantifying its a priori probability using multiple point statistics. Letting the a priori information be available in, for instance, a training image, the FM method solves an inverse problem by computing a model that satisfies not only the relation from Eq. 1 but a model that is also similar to the training image. The latter ensures that the model will be geologically reasonable.

2.1 The Maximum A Posteriori Model

Tarantola and Valette (1982) derived a probabilistic approach to solve inverse problems where the solution to the inverse problem is given by a probability density function, denoted the a posteriori distribution. This approach makes use of a prior distribution and a likelihood function to assign probabilities to all possible models. The a priori probability density function ρ describes the data independent prior knowledge of the model parameters; in the FM method we choose to define it as follows

$$\rho(\mathbf{m}) = \text{const.} \exp\big(-\alpha\ f(\mathbf{m})\big),$$

where α acts as a weighting parameter and f is a dissimilarity function presented in Sect. 2.4. Traditionally, f measures the distance between the model and an a priori model. The idea behind the FM method is the same, except we wish not to compare models directly but to compare the multiple point statistics of models. We therefore choose a traditional prior but replace the distance function such that instead of measuring the distance between models directly, we measure the dissimilarity between them. The dissimilarity is expressed as a distance between their multiple point statistics.

The likelihood function L is a probabilistic measure of how well data associated with a certain model matches the observed data, accounting for the uncertainties of the observed data,

$$L\big(\mathbf{m}, \mathbf{d}^{\text{obs}}\big) = \text{const.} \exp\left(-\frac{1}{2}\big\|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\big\|^2_{\mathbf{C}_\text{d}}\right).$$

Here, $\mathbf{C}_\text{d}$ is the data covariance matrix and the measurement errors are assumed to be independent and Gaussian distributed with mean values 0. The a posteriori distribution is then proportional to the product of the prior distribution and the likelihood

$$\sigma(\mathbf{m}) = \text{const.}\rho(\mathbf{m})L\big(\mathbf{m}, \mathbf{d}^{\text{obs}}\big).$$

The set of model parameters that maximizes the a posteriori probability density is called the maximum a posteriori (MAP) model

$$\begin{aligned}\mathbf{m}^{\text{MAP}} &= \arg\max_{\mathbf{m}}\big\{\sigma(\mathbf{m})\big\} \\ &= \arg\min_{\mathbf{m}}\big\{-\log\sigma(\mathbf{m})\big\} \\ &= \arg\min_{\mathbf{m}}\left\{\frac{1}{2}\big\|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\big\|^2_{\mathbf{C}_\text{d}} + \alpha\ f(\mathbf{m})\right\}.\end{aligned}$$

The dissimilarity function f is a measure of how well the model satisfies the a priori knowledge that is available, for example from a training image. The more similar, in some sense, the image from a set of model parameters $\mathbf{m}$ is to the training image the smaller the function value $f(\mathbf{m})$ is. Equivalently to the more traditional term $\|\mathbf{m} - \mathbf{m}^{\text{prior}}\|^2_{\mathbf{C}_\text{m}}$, stemming from a Gaussian a priori distribution of the model parameters with mean values $\mathbf{m}^{\text{prior}}$ and covariance matrix $\mathbf{C}_\text{m}$, $f(\mathbf{m})$ can be thought of as a distance. It is not a distance between $\mathbf{m}$ and the training image ($f(\mathbf{m})$ may be zero for other images than the training image), but a distance between the multiple point statistics of the image formed by the model parameters and the multiple point statistics of the training image.

2.2 The Multiple Point Statistics of an Image

Consider an image $Z = \{1, 2, \ldots, N\}$ with N voxels (or pixels if the image is only two dimensional) where the voxels can have the m different values $0, 1, \ldots, m-1$. We introduce the N variables, $z_1, z_2, \ldots, z_N$ and let z_k describe the value of the kth voxel of the image. It is assumed that the image is a realization of an unknown, random process satisfying:

1. The value of the kth voxel, z_k, is, given the values of voxels in a certain neighborhood $\mathcal{N}_k$ around voxel k, independent of voxel values not in the neighborhood. Voxel k itself is not contained in $\mathcal{N}_k$. Let $\mathbf{z}_k$ be a vector of the values of the ordered neighboring voxels in $\mathcal{N}_k$; we then have
$$f_Z(z_k|z_N, \ldots, z_{k+1}, z_{k-1}, \ldots, z_1) = f_Z(z_k|\mathbf{z}_k),$$
where f_Z denotes the conditional probability distribution of the voxel z_k given the values of the voxels within the neighborhood.
2. For an image of infinite size the geometrical shape of all neighborhoods $\mathcal{N}_k$ are identical. This implies that if voxel k has coordinates (k_x, k_y, k_z), and voxel l has coordinates (l_x, l_y, l_z), then
$$(n_x, n_y, n_z) \in \mathcal{N}_k \quad \Rightarrow \quad (n_x - k_x + l_x, n_y - k_y + l_y, n_z - k_z + l_z) \in \mathcal{N}_l.$$
3. If we assume ergodicity, that is, when two voxels, voxel k and voxel l, have the same values as their neighboring voxels, then the conditional probability distribution of voxel k and voxel l are identical
$$\mathbf{z}_k = \mathbf{z}_l \quad \Rightarrow \quad f_Z(z_k|\mathbf{z}_k) = f_Z(z_l|\mathbf{z}_l).$$

Knowing the conditionals $f_Z(z_k|\mathbf{z}_k)$ we know the multiple point statistics of the image, just as a variogram would describe the two-point statistics of an image. The basis of sequential simulation as proposed by Guardiano and Srivastava (1993) is to exploit the aforementioned assumptions to estimate the conditional probabilities $f_Z(z_k|\mathbf{z}_k)$ based on the marginals obtained from the training image, and then to use the conditional distributions to generate new realizations of the unknown random process from which the training image is a realization. The FM method, on the other hand, operates by characterizing images by their frequency distributions. As described in the following section, the frequency distribution of voxel values within the given neighborhood of an image is given by its marginal distributions. This means

that comparison of images is done by comparing their marginals. For now, the training image is assumed to be stationary. With the current formulation of the frequency distributions this is the only feasible approach. Discussion of how to avoid the assumption of stationarity exists in literature, see for instance the recent Honarkhah (2011). Some of these approaches mentioned here might also be useful for the FM method, but we will leave this to future research to determine.

2.3 Characterizing Images by their Frequency Distribution

Before presenting the FM method we define what we denote the frequency distribution. Given an image with the set of voxels $Z = \{1, \ldots, N\}$ and voxel values $z_1, \ldots, z_N$ we define the template function Ω as a function that takes as argument a voxel k and returns the set of voxels belonging to the neighborhood $\mathcal{N}_k$ of voxel k. In the FM method, the neighborhood of a voxel is indirectly given by the statistical properties of the image itself; however, the shape of a neighborhood satisfying the assumptions from Sect. 2.2 is unknown. For each training image one must therefore define a template function Ω that seeks to correctly describe the neighborhood. The choice of template function determines if a voxel is considered to be an inner voxel. An inner voxel is a voxel with the maximal neighborhood size, and the set of inner voxels, Z_{in}, of the image is therefore defined as

$$Z_{\text{in}} = \left\{ k \in Z\colon\ |\mathcal{N}_k| = \max_{l \in Z} |\mathcal{N}_l| \right\},$$

where $|\mathcal{N}_k|$ denotes the number of voxels in $\mathcal{N}_k$. Let n denote the number of voxels in the neighborhood of an inner voxel. Typically, voxels on the boundary or close to the boundary of an image will not be inner voxels. To each inner voxel z_k we assign a pattern value p_k; we say the inner voxel is the center voxel of a pattern. This pattern value is a unique identifier of the pattern and may be chosen arbitrarily. The most obvious choice is perhaps a vector value with the discrete variables in the pattern, or a scalar value calculated based on the values of the variables. The choice should be made in consideration of the implementation of the FM method. The pattern value is uniquely determined by the value of the voxel z_k and the values of the voxels in its neighborhood, $\mathbf{z}_k$. As the pattern value is determined by the values of $n + 1$ voxels, which can each have m different values, the maximum number of different patterns is m^{n+1}.

Let π_i, for $i = 1, \ldots, m^{n+1}$, count the number of patterns that have the ith pattern value. The frequency distribution is then defined as $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = [\pi_1, \ldots, \pi_{m^{n+1}}].$$

Let p_Ω denote the mapping from voxel values of an image Z to its frequency distribution $\boldsymbol{\pi}$, that is, $p_\Omega(z_1, \ldots, z_N) = \boldsymbol{\pi}$.

Figure 1 shows an example of an image and the patterns it contains for the template function that defines neighborhoods as follows

$$\mathcal{N}_k = \left\{ l \in Z \setminus \{k\}\colon\ |l_x - k_x| \leq 1, |l_y - k_y| \leq 1 \right\}.$$

Recall from Sect. 2.2 that (l_x, l_y) are the coordinates of voxel l in this two-dimensional example image. We note that for a given template function the frequency

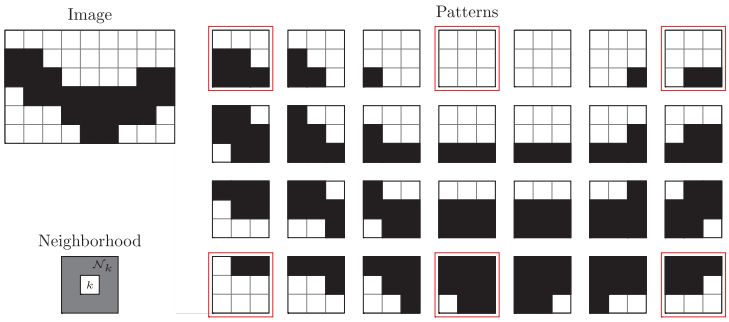


Fig. 1 Example of patterns found in an image. Notice how the image is completely described by the (ordered) patterns in every third row and column; the patterns are marked in red

distribution of an image is uniquely determined. The opposite, however, does not hold. Different images can, excluding symmetries, have the same frequency distribution. This is what the FM method seeks to exploit by using the frequency distribution to generate new images, at the same time similar to, and different from, our training image.

2.4 Computing the Similarity of Two Images

The FM method compares a solution image to a training image by comparing its frequency distribution to the frequency distribution of the training image. How dissimilar the solution image is to the training image is determined by a dissimilarity function, which assigns a distance between their frequency distributions. This distance reflects how likely the solution image is to be a realization of the same unknown process as the training image is a realization of. The bigger the distance, the more dissimilar are the frequency distributions and thereby also the images, and the less likely is the image to be a realization of the same random process as the training image. The dissimilarity function can therefore be used to determine which of two images is most likely to be a realization of the same random process as the training image is a realization of.

The dissimilarity function is not uniquely given but an obvious choice is the χ^2 distance also described in Sheskin (2004). It is used to measure the distance between two frequency distributions by measuring how similar the proportions of patterns in the frequency distributions are. Given two frequency distributions, the χ^2 distance estimates the underlying distribution. It then computes the distance between the two frequency distributions by computing each of their distances to the underlying distribution. Those distances are computed using a weighted Euclidean norm where the weights are the inverse of the counts of the underlying distribution, see Fig. 2. In our research, using the counts of the underlying distribution turns out to be a favorable weighting of small versus big differences instead of using a traditional p-norm as used by Peredo and Ortiz (2010).

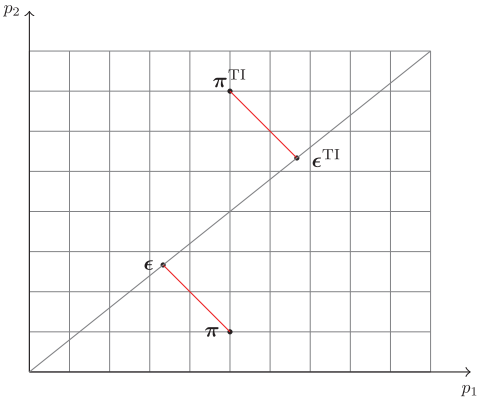


Fig. 2 Illustration of the χ^2 distance between two frequency distributions $\boldsymbol{\pi}$ and $\boldsymbol{\pi}^{\mathrm{TI}}$, each containing the counts of two different pattern values, p_1 and p_2. The difference between the frequency distributions is computed as the sum of the length of the two red line segments. The length of each line segment is computed using a weighted Euclidean norm. The counts of the underlying distribution are found as the orthogonal projection of the frequency distributions onto the a line going through the origin such that $\|\boldsymbol{\pi} - \boldsymbol{\epsilon}\|_2 = \|\boldsymbol{\pi}^{\mathrm{TI}} - \boldsymbol{\epsilon}^{\mathrm{TI}}\|_2$

Hence, given the frequency distributions of an image, $\boldsymbol{\pi}$, and of a training image, $\boldsymbol{\pi}^{\mathrm{TI}}$, and by letting

$$I = \{i \in \{1, \ldots, m^{n+1}\} \colon \pi_i^{\mathrm{TI}} > 0\} \cup \{i \in \{1, \ldots, m^{n+1}\} \colon \pi_i > 0\}, \tag{2}$$

we compute what we define as the dissimilarity function value of the image

$$c(\boldsymbol{\pi}) = \chi^2(\boldsymbol{\pi}, \boldsymbol{\pi}^{\mathrm{TI}}) = \sum_{i \in I} \frac{(\pi_i^{\mathrm{TI}} - \epsilon_i^{\mathrm{TI}})^2}{\epsilon_i^{\mathrm{TI}}} + \sum_{i \in I} \frac{(\pi_i - \epsilon_i)^2}{\epsilon_i}, \tag{3}$$

where ϵ_i denotes the counts of the underlying distribution of patterns with the ith pattern value for images of the same size as the image and ϵ_i^{TI} denotes the counts of the underlying distribution of patterns with the ith pattern value for images of the same size as the training image. These counts are computed as

$$\epsilon_i = \frac{\pi_i + \pi_i^{\mathrm{TI}}}{n_Z + n_{\mathrm{TI}}} n_Z, \tag{4}$$

$$\epsilon_i^{\mathrm{TI}} = \frac{\pi_i + \pi_i^{\mathrm{TI}}}{n_Z + n_{\mathrm{TI}}} n_{\mathrm{TI}}, \tag{5}$$

where n_Z and n_{TI} are the total number of counts of patterns in the frequency distributions of the image and the training image, that is, the number of inner voxels in the image and the training image, respectively.

2.5 Solving Inverse Problems

We define the frequency matching method for solving inverse problems formulated as least squares problems using geologically complex a priori information as the fol-

lowing optimization problem

$$\begin{aligned} &\min_{z_1,\ldots,z_N} \left\| \mathbf{d}^{\text{obs}} - g(z_1,\ldots,z_N) \right\|^2_{\mathbf{C}_\text{d}} + \alpha\, c(\boldsymbol{\pi}), \\ &\text{w.r.t. } \boldsymbol{\pi} = p_\Omega(z_1,\ldots,z_N), \\ &z_k \in \{0,\ldots,m-1\} \quad \text{for } k = 1,\ldots,N, \end{aligned} \tag{6}$$

where $c(\boldsymbol{\pi})$ is the dissimilarity function value of the solution image defined by Eq. 3 and α is a weighting parameter. The forward operator g, which traditionally is a mapping from model space to data space, also contains the mapping of the categorical values $z_k \in \{0,\ldots,m-1\}$ for $k = 1,\ldots,N$ of the image into the model parameters $\mathbf{m}$ that can take m different discrete values.

The value of α cannot be theoretically determined. It is expected to depend on the problem at hand; among other factors its resolution, the chosen neighborhood function and the dimension of the data space. It can be thought of as playing the same role for the dissimilarity function as the covariance matrix $\mathbf{C}_\text{d}$ does for the data misfit. So it should in some sense reflect the variance of the dissimilarity function and in that way determine how much trust we put in the dissimilarity value. Variance, or trust, in a training image is difficult to quantify, as the training image is typically given by a geologist to reflect certain expectations to model. Not having a theoretical expression for α therefore allows us to manipulate the α value to loosely quantify the trust we have in the training image. In the case where we have accurate data but only a vague idea of the structures of the subsurface the α can be chosen low, in order to emphasize the trust we have in the data and the uncertainty we have of the structure of the model. In the opposite case, where data are inaccurate but the training image is considered to be a very good description of the subsurface, the α value can be chosen high, to give the dissimilarity function more weight.

Due to the typically high number of model parameters, the combinatorial optimization problem should be solved by use of an iterative solution method; such a method will iterate through the model space and search for the optimal solution. While the choice of solution method is less interesting when formulating the FM method, it is of great importance when applying it. The choice of solution method and the definition of how it iterates through the solution space by perturbing images has a significant impact on the feasibility of the method in terms of its running time. As we are not sampling the solution space we do not need to ensure that the method captures the uncertainty of the model parameters, and the ideal would be a method that converges directly to the maximum a posteriori solution. While continuous optimization problems hold information about the gradient of the objective function that the solution method can use to converge to a stationary solution, this is not the case for our discrete problem. Instead we consider the multiple point statistics of the training image when perturbing a current image and in that way we seek to generate models which better match the multiple point statistics of the training image and thus guide the solution method to the maximum a posteriori model.

2.6 Properties of the Frequency Matching Method

The FM method is a general method and in theory it can be used to simulate any type of structure, as long as a valid training image is available and a feasible template

function is chosen appropriately. If neighborhoods are chosen too small, the method will still be able to match the frequency distributions. However, it will not reproduce the spatial structures simply because these are not correctly described by the chosen multiple point statistics and as a result the computed model will not be realistic. If neighborhoods are chosen too big, CPU cost and memory demand will increase, and as a result the running time per iteration of the chosen solution method will increase. Depending on the choice of iterative solution method, increasing the size n of the neighborhood is likely to also increase the number of iterations needed and thereby increase the convergence time. When the size of neighborhoods is increased, the maximum number of different patterns, m^{n+1}, is also increased. The number of different patterns present is, naturally, limited by the number of inner voxels, which is significantly smaller than m^{n+1}. In fact, the number of patterns present in an image is restricted further as training images are chosen such that they describe a certain structure. This structure is also sought to be described in the solutions. The structure is created by repetition of patterns, and the frequency distributions will reveal this repetition by having multiple counts of the same pattern. This means, the number of patterns with non-zero frequency is greatly smaller than m^{n+1} resulting in the frequency distributions becoming extremely sparse. For bigger test cases, with millions of parameters, patterns consisting of hundreds of voxels and multiple categories, this behavior needs to be investigated further.

The dimension of the images, if they are two or three dimensional, is not important to the FM method. The complexity of the method is given by the maximal size of neighborhoods, n. The increase in n as a result of going from two- to three-dimensional images is therefore more important than the actual increase in physical dimensions. In fact, when it comes to assigning pattern values a neighborhood is, regardless of its physical dimension, considered one dimensional where the ordering of the voxels is the important aspect. Additionally, the number of categories of voxel values m does not influence the running time per iteration. As with the number of neighbors, n, it only influences the number of different possible patterns m^{n+1} and thereby influences the sparsity of the frequency distribution of the training image. The higher m is, the sparser is the frequency distribution. It is expected that the sparsity of the frequency distribution affects the level of difficulty of the combinatorial optimization problem.

Strebelle (2002) recommends choosing a training image that is at least twice as large as the structures it describes; one must assume this advice also applies to the FM method. Like the *snesim* algorithm, the FM method can approximate continuous properties by discretizing them into a small number of categories. One of the advantages of the FM method is that by matching the frequency distributions it indirectly ensures that the proportion of voxels in each of the m categories is consistent between the training image and the solution image. It is therefore not necessary to explicitly account for this ratio. Unlike the *snesim* algorithm, the computed solution images therefore need very little post treatment—in the current implementation the solution receives no post treatment. However, the α parameter does allow for the user to specify how strictly the frequency distributions should be matched. In the case where the data are considered very informative or the training image is considered far from reality, decreasing the α allows for the data to be given more weight and the multiple point statistics will not be as strictly enforced.

Constraints on the model parameters can easily be dealt with by reducing the feasible set $\{0, \ldots, m-1\}$ for those values of k in the constraints of the problem stated in Eq. 6. The constrained voxels remain part of the image Z and when computing the frequency distribution of an image they are not distinguished from non-constrained voxels. However, when perturbing an image all constraints of the inverse problem should at all times be satisfied and conditioned to the hard data. The additional constraints on the model parameters will therefore be honored.

3 Implementation

This section describes the current implementation of the frequency matching method. Algorithm 1 gives a general outline of how to apply the FM method, that is, how to solve the optimization problem from Eq. 6 with an iterative optimization method. In the remainder of the section, the implementation of the different parts of the FM method will be discussed. It should be noted that the implementation of the FM method is not unique; for instance, there are many options for how the solution method iterates through the model space by perturbing models. The different choices should be made depending on the problem at hand and the current implementation might not be favorable for some given problems. The overall structure in Algorithm 1 will be valid regardless of what choices are made on a more detailed level.

Algorithm 1: The Frequency Matching Method

Input: Training image, Z^{TI}, Starting image Z
Output: Maximum a posteriori image Z^{FM}
Compute frequency distribution of training image $\boldsymbol{\pi}^{\mathrm{TI}}$ and pattern list $\mathbf{p}$ (Algorithm 2)
Compute partial frequency distribution of starting image $\boldsymbol{\pi}$ (Algorithm 3)
while *not converged* **do**
  Compute perturbed image $\overline{Z}$ based on Z (Algorithm 4)
  Compute partial frequency distribution of perturbed image $\overline{\boldsymbol{\pi}}$ (Algorithm 5)
  if *accept the perturbed image* **then**
    Set $Z \leftarrow \overline{Z}$ and $\boldsymbol{\pi} \leftarrow \overline{\boldsymbol{\pi}}$
  end
end

The current implementation is based on a Simulated Annealing scheme. Simulated Annealing is a well-known heuristic optimization method first presented by Kirkpatrick et al. (1983) as a solution method for combinatorial optimization problems. The acceptance of perturbed images is done using an exponential cooling rate and the parameters controlling the cooling are tuned to achieve an acceptance ratio of approximately 15 accepted perturbed models for each 100 suggested perturbed models. A perturbed model is generated by erasing the values of the voxels in a part of the image and then re-simulating the voxel values by use of sequential simulation.

3.1 Reformulation of the Dissimilarity Function

The definition of the dissimilarity function from Eq. 3 has one great advantage that we for computational reasons simply cannot afford to overlook. As discussed previously, the frequency distributions are expected to be sparse as the number of patterns present in an image is significantly smaller than m^{n+1}. This means that a lot of the terms in the dissimilarity function from Eq. 3 will be zero, yet the dissimilarity function can be simplified further. It will be shown that the dissimilarity function value of a frequency distribution, $c(\boldsymbol{\pi})$, given the frequency distribution of a training image, $\boldsymbol{\pi}$, can be computed using only entries of $\boldsymbol{\pi}$ where $\boldsymbol{\pi}^{\mathrm{TI}} > 0$. In other words, to compute the dissimilarity function value of an image we need only to know the count of patterns in the image that also appear in the training image. Computationally, this is a great advantage as we can disregard the patterns in our solution image that do not appear in the training image and we need not compute nor store the entire frequency distribution of our solution image, which is shown by inserting the expressions of the counts for the underlying distribution defined by Eqs. 4 and 5

$$\begin{aligned} c(\boldsymbol{\pi}) &= \sum_{i\in I} \frac{(\pi_i^{\mathrm{TI}} - \epsilon_i^{\mathrm{TI}})^2}{\epsilon_i^{\mathrm{TI}}} + \sum_{i\in I} \frac{(\pi_i - \epsilon_i)^2}{\epsilon_i} \\ &= \sum_{i\in I} \frac{(\sqrt{\frac{n_Z}{n_{\mathrm{TI}}}}\,\pi_i^{\mathrm{TI}} - \sqrt{\frac{n_{\mathrm{TI}}}{n_Z}}\pi_i)^2}{\pi_i^{\mathrm{TI}} + \pi_i}. \end{aligned} \tag{7}$$

This leads to the introduction of the following two subsets of I

$$\begin{aligned} I_1 &= \left\{ i \in I : \pi_i^{\mathrm{TI}} > 0 \right\}, \\ I_2 &= \left\{ i \in I : \pi_i^{\mathrm{TI}} = 0 \right\}. \end{aligned}$$

The two subsets form a partition of I as they satisfy $I_1 \cup I_2 = I$ and $I_1 \cap I_2 = \emptyset$. The dissimilarity function Eq. 7 can then be written as

$$\begin{aligned} c(\boldsymbol{\pi}) &= \sum_{i\in I_1} \frac{(\sqrt{\frac{n_Z}{n_{\mathrm{TI}}}}\pi_i^{\mathrm{TI}} - \sqrt{\frac{n_{\mathrm{TI}}}{n_Z}}\pi_i)^2}{\pi_i^{\mathrm{TI}} + \pi_i} + \frac{n_{\mathrm{TI}}}{n_Z} \sum_{i\in I_2} \pi_i \\ &= \sum_{i\in I_1} \frac{(\sqrt{\frac{n_Z}{n_{\mathrm{TI}}}}\pi_i^{\mathrm{TI}} - \sqrt{\frac{n_{\mathrm{TI}}}{n_Z}}\pi_i)^2}{\pi_i^{\mathrm{TI}} + \pi_i} + \frac{n_{\mathrm{TI}}}{n_Z}\left(n_Z - \sum_{i\in I_1} \pi_i \right) \end{aligned} \tag{8}$$

recalling that $\sum_{i\in I} \pi_i = n_Z$ and that $\pi_i = 0$ for $i \notin I$.

A clear advantage of this formulation of the dissimilarity function is that the entire frequency distribution $\boldsymbol{\pi}$ of the image does not need to be known; as previously stated, it only requires the counts π_i of the patterns also found in the training image, which is for $i \in I_1$.

3.2 Computing and Storing the Frequency Distributions

The formulation of the dissimilarity function from Eq. 3 and later Eq. 8 means that it is only necessary to store non-zero entries in a frequency distribution of a training

image $\boldsymbol{\pi}^{\mathrm{TI}}$. Algorithm 2 shows how the frequency distribution of a training image is computed such that zero entries are avoided. The algorithm also returns a list $\mathbf{p}$ with the same number of elements as the frequency distribution and it holds the pattern values corresponding to each entry of $\boldsymbol{\pi}^{\mathrm{TI}}$.

Algorithm 2: Frequency Distribution of a Training Image

```
Input: Training Image Z^TI
Output: Frequency distribution π^TI, list of pattern values p
Initialization: empty list π^TI, empty list p
for each inner voxel, i.e., k ∈ Z^TI_in do
    Extract pattern k
    Compute pattern value p_k
    if the pattern was previously found then
        Add 1 to the corresponding entry of π^TI
    else
        Add p_k to the list of pattern values p
        Set the corresponding new entry of π^TI equal to 1
    end
end
```

Algorithm 3 computes the partial frequency distribution $\boldsymbol{\pi}$ of an image that is needed to evaluate the dissimilarity function $c(\boldsymbol{\pi}) = \chi^2(\boldsymbol{\pi}, \boldsymbol{\pi}^{\mathrm{TI}})$ from Eq. 8. The partial frequency distribution only stores the frequencies of the patterns also found in the training image.

Algorithm 3: Partial Frequency Distribution of an Image

```
Input: Image Z, list of pattern values p from the training image
Output: Partial frequency distribution π
Initialization: all zero list π (same length as p)
for each inner voxel, i.e., k ∈ Z_in do
    Extract pattern k
    Compute pattern value p_k
    if the pattern is found in the training image then
        Add 1 to the corresponding entry of π
    end
end
```

3.3 Perturbation of an Image

The iterative solver moves through the model space by perturbing models and this is the part of the iterative solver that leaves the most choices to be made. An intuitive but naive approach would be to simply change the value of a random voxel. This will result in a perturbed model that is very close to the original model, and it will therefore require a lot of iterations to converge. The current implementation changes the values of a block of voxels in a random place of the image.

Before explaining in detail how the perturbation is done, let $Z^{\text{cond}} \subset Z$ be the set of voxels that we have hard data for, which means their value is known and should be conditioned to. First a voxel k is chosen randomly. Then the value of all voxels in a domain $\mathcal{D}_k \subset (Z \setminus Z^{\text{cond}})$ around voxel k are erased. Last, the values of the voxels in $\mathcal{D}_k$ are simulated using sequential simulation. The size of the domain should be chosen to reflect how different the perturbed image should be from the current image. The bigger the domain, the fewer iterations we will expect the solver will need to iterate through the model space to converge, but the more expensive an iteration will become. Choosing the size of the domain is therefore a trade-off between number of iterations and thereby forward calculations and the cost of computing a perturbed image.

Algorithm 4 shows how an image is perturbed to generate a new image.

Algorithm 4: Perturbation of an Image

Input: Image Z, partial frequency distribution $\boldsymbol{\pi}$ of Z
Output: Perturbed image $\overline{Z}$
Initialization: set $\overline{\boldsymbol{\pi}} = \boldsymbol{\pi}$
Pick random voxel k
for *each voxel l around voxel k, i.e., $l \in \mathcal{D}_k$* **do**
  Erase the value of voxel l, i.e., z_l is unassigned
end
for *each unassigned voxel l around voxel k, i.e., $l \in \mathcal{D}_k$* **do**
  Simulate z_l given all assigned voxels in $\mathcal{N}_l$.
end

3.4 Updating the Frequency Distribution

As a new image is created by changing the value of a minority of the voxels, it would be time consuming to compute the frequency distribution of all voxel values of the new image when the frequency distribution of the old image is known. Recall that n is the maximum number of neighbors a voxel can have; inner voxels have exactly n neighbors. Therefore, in addiction to changing its own pattern value, changing the value of a voxel will affect the pattern value of at most n other voxels. This means that we obtain the frequency distribution of the new image by performing at most $n+1$ subtractions and $n+1$ additions per changed voxel to the entries of the already known frequency distribution.

The total number of subtractions and additions can be lowered further by exploiting the block structure of the set of voxels perturbed. The pattern value of a voxel will be changed when any of its neighboring voxels are perturbed, but the frequency distribution need only be updated twice for each affected voxel. We introduce a set of voxels Z^{aff}, which is the set of voxels who are affected when perturbing image Z into $\overline{Z}$, that is, the set of voxels whose pattern values are changed when perturbing image Z into image $\overline{Z}$

$$Z^{\text{aff}} = \{k \in Z:\ p_k \neq \overline{p}_k\}. \tag{9}$$

How the partial frequency distribution is updated when an image is perturbed is illustrated in Algorithm 5.

Algorithm 5: Update Partial Frequency Distribution of an Image

Input: Image Z, partial frequency distribution $\boldsymbol{\pi}$ of Z, perturbed image $\overline{Z}$, set of affected voxels Z^{aff}, set of pattern values $\mathbf{p}$ from the training image
Output: Partial frequency distribution $\overline{\boldsymbol{\pi}}$ of $\overline{Z}$
Initialization: set $\overline{\boldsymbol{\pi}} = \boldsymbol{\pi}$
for *each affected voxel, i.e.,* $k \in Z^{\text{aff}}$ **do**
  Extract pattern k from both Z and $\overline{Z}$
  Compute both pattern values p_k and $\overline{p}_k$
  if *the pattern* p_k *is present in the training image* **then**
    Subtract 1 from the corresponding entry of $\overline{\boldsymbol{\pi}}$
  end
  if *the pattern* $\overline{p}_k$ *is present in the training image* **then**
    Add 1 to the corresponding entry of $\overline{\boldsymbol{\pi}}$
  end
end

As seen in Algorithm 1, the FM method requires in total two computations of a frequency distribution, one for the training image and one for the initial image. The FM method requires one update of the partial frequency distribution per iteration. As the set of affected voxels Z^{aff} is expected to be much smaller than the total image Z, updating the partial frequency distribution will typically be much faster than recomputing the entire partial frequency distribution even for iterations that involve changing the values of a large set of voxels.

3.5 Multigrids

The multigrid approach from Strebelle (2002) that is based on the concept initially proposed by Gómez-Hernández (1991) and further developed by Tran (1994) can also be applied in the FM method. Coarsening the images allows the capture of large-scale structures with relatively small templates. As in the *snesim* algorithm, the results from a coarse image can be used to condition upon for a higher resolution image.

The multigrid approach is applied by running the FM method from Algorithm 1 multiple times. First, the algorithm is run on the coarsest level. Then the resulting image, with increased resolution, is used as a starting image on the next finer level, and so on. The resolution of an image can be increased by nearest neighbor interpolation.

4 Example: Crosshole Tomography

Seismic borehole tomography involves the measurement of seismic travel times between two or more boreholes in order to determine an image of seismic velocities in the intervening subsurface. Seismic energy is released from sources located in one borehole and recorded at multiple receiver locations in another borehole. In this way a dense tomographic data set that covers the interborehole region is obtained.

Consider a setup with two boreholes. The horizontal distance between them is ΔX and they both have the depth ΔZ. In each borehole a series of receivers and sources

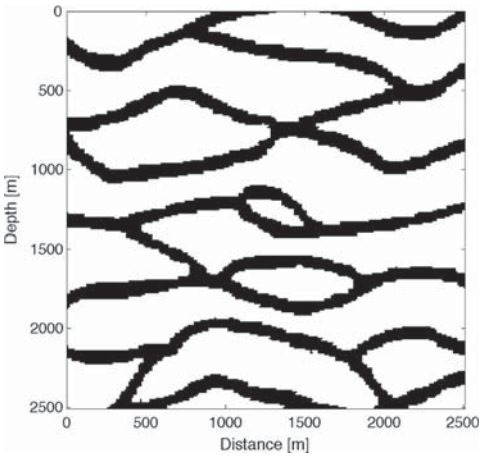


Fig. 3 Training image (resolution: 251 × 251 pixels)

Table 1 Parameter values for the test case

ΔX	500 m
ΔZ	1,200 m
Δx	10 m
Δz	10 m
d_{s}	250 m
d_{r}	100 m
v_{low}	1,600 m/s
v_{high}	2,000 m/s

is placed. The vertical domain between the two boreholes is divided into cells of dimensions Δx by Δz and it is assumed that the seismic velocity is constant within each cell. The model parameters of the problem are the propagation speeds of each cell. The observed data are the first arrival times of the seismic signals. For the series of sources and receivers in each borehole the distances between the sources are d_s and the distances between the receivers are d_r. We assume a linear relation between the data (first arrival times) and the model (propagation speed) from Eq. 1. The sensitivity of seismic signals is simulated as straight rays. However, any linear sensitivity kernel obtained using, for example, curvilinear rays or Fresnel zone-based sensitivity, can be used.

It is assumed that the domain consists of zones with two different propagation speeds, v_{low} and v_{high}. Furthermore a horizontal channel structure of the zones with high propagation speed is assumed. Figure 3 shows the chosen training image with resolution 251 cells by 251 cells where each cell is Δx by Δz. The training image is chosen to express the a priori information about the model parameters. The background (white pixels) represents a low velocity zone and the channel structures (black

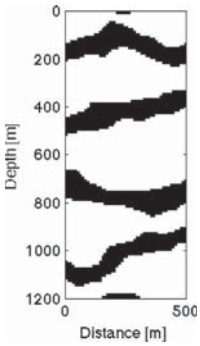


Fig. 4 Reference model (resolution: 50×120 pixels)

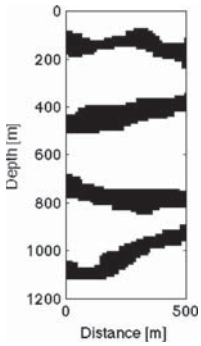


Fig. 5 Computed model for $\alpha = 1.8 \times 10^{-2}$ (resolution: 50×120 pixels)

pixels) are the high velocity zones. The problem is scalable and for the example we have chosen the parameters presented by Table 1.

The template function is chosen, such that the neighborhood of pixel k is the following set of pixels

$$\mathcal{N}_k = \{l \in Z \setminus \{k\}\colon |l_x - k_x| \leq 4, |l_z - k_z| \leq 3\}.$$

Recall that pixel l has the coordinates (l_x, l_z); the first coordinate being the horizontal distance from the left borehole and the second coordinate being the depth, both measured in pixels. To compute a perturbed image, the domain used in Algorithm 4 is defined as follows

$$\mathcal{D}_k = \{l \in Z \setminus Z^{\text{cond}}\colon |l_x - k_x| \leq 7, |l_z - k_z| \leq 7\}.$$

The values of all pixels $l \in \mathcal{D}_k$ will be re-simulated using Sequential Simulation conditioned to the remaining pixels $l \notin \mathcal{D}_k$. We are not using any hard data in the example, which means $Z^{\text{cond}} = \emptyset$.

This choice of template function yields $n = 34$ where the geometrical shape of the neighborhood of inner pixels is a 7 pixels by 5 pixels rectangle. This is chosen based

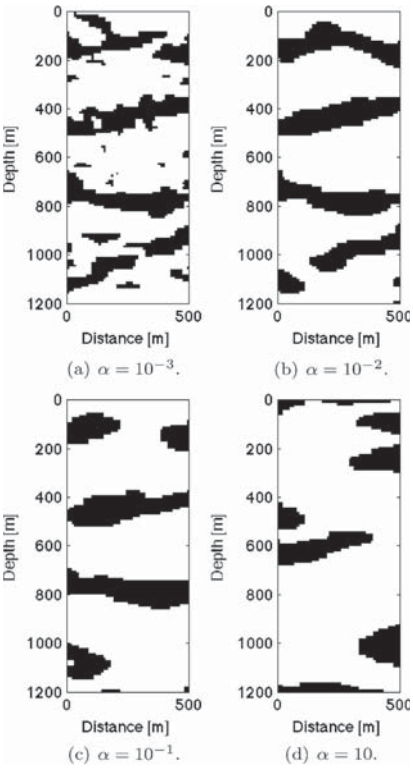


Fig. 6 The computed models for increasing values of α: (**a**) $\alpha = 10^{-3}$, (**b**) $\alpha = 10^{-2}$, (**c**) $\alpha = 10^{-1}$, (**d**) $\alpha = 10$

on the trends in the training image, where the distance of continuity is larger horizontally than vertically. However, it should be noted that this choice of template function is not expected to meet the assumptions of conditional independence of Sect. 2.2. The distance of continuity in the training image appears much larger horizontally than only seven pixels, and vertically the width of the channels is approximately ten pixels. This implies that, despite matched frequency distributions, a computed solution will not necessarily be recognized to have the same visual structures as the training image. The goal is solve the inverse problem which involves fitting the data and therefore, as our example will show, neighborhoods of this size are sufficient. The data-fitting term of the objective function guides the solution method, such that the structures from the training image are correctly reproduced. The low number of neighbors constrains the small-scale variations, which are not well-determined by the travel time data. However, the travel time data successfully determine the large-scale structures. The template function does not need to describe structures of the largest scales of the training image as long as the observed data are of a certain quality.

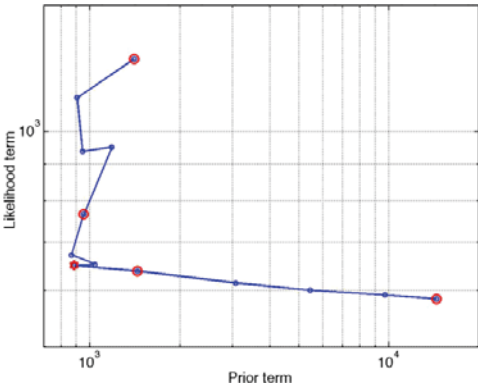


Fig. 7 L-curve used to determine the optimal α value. Models have been computed for 13 logarithmically distributed values of α ranging from 1 (*upper left corner*) to 10^{-3} (*lower right corner*). Each of the 13 models is marked with a blue circle. See the text for further explanation

Figure 4 shows the reference model that describes what is considered to be the true velocity profile between the two boreholes. The image has been generated by the *snesim* algorithm (Strebelle 2002) using the multiple point statistics of the training image. The arrival times $\mathbf{d}$ for the reference model $\mathbf{m}^{\text{ref}}$ are computed by a forward computation, $\mathbf{d} = G\mathbf{m}^{\text{ref}}$. We define the observed arrival times $\mathbf{d}^{\text{obs}}$ as the computed arrival times $\mathbf{d}$ added 5 % Gaussian noise. Figure 5 shows the solution computed using 15,000 iterations for $\alpha = 1.8 \times 10^{-2}$. The solution resembles the reference model to a high degree. The FM method detected the four channels; their location, width and curvature correspond to the reference model. The computations took approximately 33 minutes on a Macbook Pro 2.66 GHz Intel Core 2 Duo with 4 GB RAM.

Before elaborating on how the α value was determined, we present some of the models computed for different values of α. Figure 6 shows the computed models for four logarithmically distributed values of α between 10^{-3} and 10^{1}. It is seen how the model for lowest value of α is geologically unrealistic and does not reproduce the a priori expected structures from the training image as it primarily is a solution to the ill-posed, under-determined, data-fitting problem. As α increases, the channel structures of the training image are recognized in the computed models. However, for too large α values the solutions are dominated by the χ^2 term as the data have been deprioritized, and the solutions are not geologically reasonable either. As discussed, the chosen template is too small to satisfy the conditions from Sect. 2.2, yielding models that do in fact minimize the χ^2 distance, but do not reproduce the structures form the training image. The data misfit is now assigned too little weight to help compensate for the small neighborhoods, and the compromise between minimizing the data misfit and minimizing the dissimilarity that before worked out well is no longer present.

We propose to use the L-curve method (Hansen and O'Leary 1993) to determine an appropriate value of α. Figure 7 shows the value of $\chi^2(\mathbf{m}^{\text{FM}})$ versus the value of $\frac{1}{2}\|g(\mathbf{m}^{\text{FM}}) - \mathbf{d}^{\text{obs}}\|^2_{\mathbf{C}_{\text{d}}}$ for 13 models. The models have been computed for logarithmically distributed values of α ranging from 1 (upper left corner) to 10^{-3} (lower right corner). Each of the 13 models is marked with a blue circle. The models from Fig. 6

are furthermore marked with a red circle. The model from Fig. 5 is marked with a red star. We recognize the characteristic L-shaped behavior in the figure and the model from Fig. 5 is the model located in the corner of the L-curve. The corresponding value $\alpha = 1.8 \times 10^{-2}$ is therefore considered an appropriate value of α.

5 Conclusions

We have proposed the frequency matching method which enables us to quantify a probability density function that describes the multiple point statistics of an image. In this way, the maximum a posteriori solution to an inverse problem using training image-based complex prior information can be computed. The frequency matching method formulates a closed form expression for the a priori probability of a given model. This is obtained by comparing the multiple point statistics of the model to the multiple point statistics from a training image using a χ^2 dissimilarity distance.

Through a synthetic test case from crosshole tomography, we have demonstrated how the frequency matching method can be used to determine the maximum a posteriori solution. When the a priori distribution is used in inversion, a parameter α is required. We have shown how we are able to recreate the reference model by choosing this weighing parameter appropriately. Future work could focus on determining the theoretically optimal value of α as an alternative to using the L-curve method.

Acknowledgements The present work was sponsored by the Danish Council for Independent Research —Technology and Production Sciences (FTP grant no. 274-09-0332) and DONG Energy.

References

Arpat GB (2005) Sequential simulation with patterns. PhD thesis, Stanford University

Caers J, Hoffman T (2006) The probability perturbation method: a new look at Bayesian inverse modeling. Math Geol 38:81–100

Caers J, Zhang T (2004) Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. In: Grammer M, Harris PM, Eberli GP (eds) Integration of outcrop and modern analogs in reservoir modeling, AAPG Memoir 80, AAPG, Tulsa, pp 383–394

Cressie N, Davidson J (1998) Image analysis with partially ordered Markov models. Comput Stat Data Anal 29(1):1–26

Gómez-Hernández JJ (1991) A stochastic approach to the simulation of block conductivity fields conditioned upon data measured at a smaller scale. PhD thesis, Stanford University

Guardiano F, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate moments. In: Geostatics-Troia, vol 1. Kluwer Academic, Dordrecht, pp 133–144

Hansen PC, O'Leary DP (1993) The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J Sci Comput 14:1487–1503

Hansen TM, Cordua KS, Mosegaard K (2008) Using geostatistics to describe complex a priori information for inverse problems. In: Proceedings Geostats 2008. pp 329–338

Hansen TM, Cordua KS, Mosegaard K (2012) Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. Comput Geosci 16:593–611

Honarkhah M (2011) Stochastic simulation of patterns using distance-based pattern modeling. PhD dissertation, Stanford University

Jafarpour B, Khodabakhshi M (2011) A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. Math Geosci 43:133–146
Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680
Peredo O, Ortiz JM (2010) Parallel implementation of simulated annealing to reproduce multiple-point statistics. Comput Geosci 37:1110–1121
Sheskin D (2004) Handbook of parametric and nonparametric statistical procedures. Chapman & Hal/CRC, London, pp 493–500
Stien M, Kolbjørnsen O (2011) Facies modeling using a Markov mesh model specification. Math Geosci 43:611–624
Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34:1–21
Suzuki S, Caers J (2008) A distance-based prior model parameterization for constraining solutions of spatial inverse problems. Math Geosci 40:445–469
Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia
Tarantola A, Valette B (1982) Inverse problems = quest for information. J Geophys 50:159–170
Tjelmeland H, Besag J (1998) Markov random fields with higher-order interactions. Scand J Stat 25:415–433
Tran TT (1994) Improving variogram reproduction on dense simulation grids. Comput Geosci 7:1161–1168

Appendix A10:
History Matching: Towards Geologically Reasonable Models

Authors:

Yulia Melnikova, Knud Skou Cordua, and Klaus Mosegaard

Published in:

Proceeding of EAGE international conference on “Integrated Reservoir Modelling: Are we doing it right?”

Dubai, United Arab Emirates.

25 – 28 November 2012

RM25

History Matching: Towards Geologically Reasonable Models

Y. Melnikova* (Technical University of Denmark), K.S. Cordua (Technical University of Denmark) & K. Mosegaard (Technical University of Denmark)

SUMMARY

This work focuses on the development of a new method for history matching problem that through a deterministic search finds a geologically feasible solution. Complex geology is taken into account evaluating multiple point statistics from earth model prototypes - training images.
Further a function that measures similarity between statistics of a training image and statistics of any smooth model is introduced and its analytical gradient is computed. This allows us to apply any gradient-based method to history matching problem and guide a solution until it satisfies both production data and complexity of a prior model with desired accuracy. As a consequence of the approach, we sufficiently decrease the amount of forward simulations needed to resolve historical data and prior information.

Introduction

History matching is an essential part of reservoir characterization process. Reliable reservoir models must fit production history and feature expected geology. Therefore geological a priori information should be included in the estimation of reservoir parameters. Due to the high computational cost of forward simulations (reservoir simulator runs) use of Monte-Carlo techniques can be unfeasible.

A fast compromise solution would be to find an approximation of the maximum a posteriori solution (MAP). To succeed in this task the probability of the model to resolve geological features (prior probability) must be estimated. Recently Lange et al. (2012) suggested the Frequency Matching (FM) method for solving inverse problems by use of geologically realistic prior information. In the FM approach the a priori information takes the form of multiple-point statistics learned from reservoir geological prototypes - training images (e.g. Guardiano and Srivastava 1993). The attractiveness of the FM method lies in its ability to quantify the prior probability of the proposed model and hence iteratively guide the model towards the maximum a posteriori solution. The FM method solves a combinatorial optimization problem, perturbing the model in a discrete manner until it explains both production data and a priori information. In practice, this requires a lot of forward simulations and can be impractical for solving history matching problems.

While following the philosophy of the Frequency Matching method, we suggest a differentiable expression for a complex prior, so that, as a result, the approximation of the MAP solution can be found by gradient-based techniques with much fewer forward simulations required.

Methodology

We suggest a gradient-based method for obtaining geologically feasible solutions of history matching problem. The algorithm integrates production data and complex geological a priori information into a single objective function. Importantly, we propose a differentiable formulation of a priori information.

As a priori information, we use multiple point statistics derived from training images, which characterizes the expected spatial distribution of the sought physical property, for instance, permeability. Similar to Lange et al. (2012) we define an optimization problem, i.e. to minimize:

$$O(\mathbf{m}) = \frac{1}{2}\left\| d^{obs} - g(\mathbf{m}) \right\|_{C_d}^2 + f(\mathbf{m}, \mathrm{TI}) \tag{1}$$

Reservoir parameters $\mathbf{m}$ are then inferred by minimizing two misfits: 1) between observed production data d^{obs} and reservoir response $g(\mathbf{m})$ and 2) between statistics of the model (test image) $\mathbf{m}$ and statistics of the training image TI. One way to collect the statistics is to apply a scanning template to an image and compute the frequency distribution of the event defined by the chosen template (Lange et al., 2012). The result will be the histogram that describes the image uniquely. The distance (defined in some sense) between the histogram of the training image and one of the test image estimates their statistical similarity.

The challenge in the gradient-based approach is to define a differentiable similarity measure between the continuous image $\mathbf{m}$ and the discrete training image TI. Keeping the idea of histogram in mind, we first define the similarity function h_{ij} between a continuous pattern i and a discrete pattern j, using the normalized Euclidian distance d_{ij} between their pixels values:

$$h_{ij} = \frac{1}{(1 + A d_{ij}^k)^p} \tag{2}$$

Here A, k, p are adjustable parameters. Then the pseudo-histogram is constructed calculating the "contributions" h_{ij} of patterns in the image to all possible discrete patterns. The number of the histogram bins is equal to the number of all possible discrete patterns, i.e. $c^{N_{pat}}$, where c is the number of categories in the training image and N_{pat} is the number of pixels in the pattern.

In (1) the function $f(\mathbf{m}, \mathrm{TI})$ is the L2 norm of the difference between the pseudo-histograms of the training image TI and the test image $\mathbf{m}$.

For solving (1) we chose the unconstrained implementation of the LBFGS method (Zhu et al. 1997), which is known to be efficient for history matching problems (Oliver et al. 2008). To use unconstrained optimization we applied the logarithmic scaling of reservoir parameters proposed in Gao and Reynolds (2006):

$$x_i = \ln\left[\frac{m_i - m^{low}}{m^{up} - m_i}\right] \tag{3}$$

Here $i = 1,\ldots,n$, where n is the number of pixels in the test image $\mathbf{m}$, m^{low} and m^{up} are the lower and upper scaling boundaries respectively. Global criterion method (Marler and Arora 2004) was used to combine the data misfit and prior terms into one objective function. This yielded to the final look of the objective function:

$$O^*(\mathbf{m}) = \left[\frac{\frac{1}{2}\left\|d^{obs} - g(\mathbf{m})\right\|^2_{C_d} - h^*}{h^*}\right]^2 + \left[\frac{f(\mathbf{m}, \mathrm{TI}) - f^*}{f^*}\right]^2 \tag{4}$$

Here h^* and f^* are the target values for data and prior misfits respectively. For forward simulations E300 reservoir simulator was used (Schlumberger GeoQuest 2009). The gradient of the data misfit term in (4) was evaluated using the adjoint calculation implemented in E300. The gradient of the prior term in (4) was calculated analytically.

As in any gradient-based technique, solution and convergence properties of the suggested method are strongly dependent on the initial guess and quality of the production and statistical data. In case of a poor choice of the template size geological features cannot be reproduced. However, large amount of data may compensate for the lack of statistical information. In the numerical example below we will see how sufficient wells coverage yields the correct length of geological features, while a priori information resolves their width in agreement with training image.

Numerical example

In the test study we aim at reconstructing permeability field of a 2D synthetic oil reservoir of 49x49x1cells. The true permeability and wells (9 injectors, triangles and 9 producers, circles) are shown in ***Figure 1***. Training image of 200x200 pixels (***Figure 2***) has two categories and features highly permeable channels of 10000 mD and 500 mD background. Notice the scaling boundaries of 450 mD for m^{low} and 10500 mD for m^{up}. Production data were generated by running a forward simulation with the true permeability model and adding 5% of Gaussian noise. Specifically, the reservoir was in production for 210 days and the data were collected every 30 days. For history matching we used BHP values from the injectors and oil rates from the producers (126 measurements in total).

A priori information was collected applying a 1D-scanning template of 10 pixels in vertical direction. We let the template to take care about the width of the channels, while the production data assure the horizontal continuity.

While parameters A, k, p, in (2) are empirical, such values as 100, 2 and -1 respectively provide the optimal quality of the reconstructed image and may serve as a general recommendation.

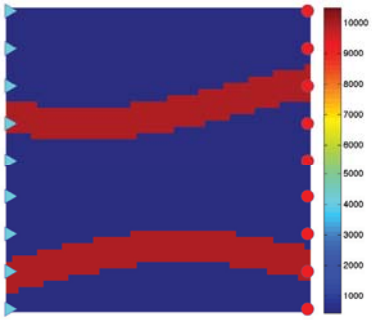


Figure 1 *True permeability model, 49x49 pixels.*

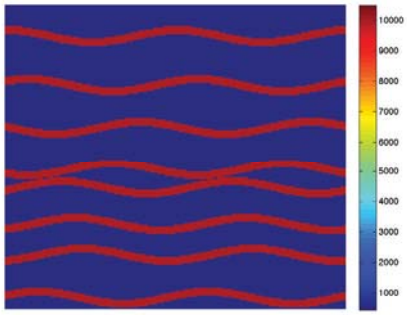


Figure 2 *Training image, 200 x200 pixels .*

The initial model (see ***Figure 3***) gives the data misfit of the order of 10^5 and the histograms misfit - of 10^{-1}. In the optimization framework given by (4), we set the target values as 20 and 0.005 for the data and the histograms misfits respectively.

Figure 4 shows the solution at the 97th iteration. Visual inspection tells us that geological features were successfully reproduced. Additionally, the expected order of 10^{-3} in the histograms misfit was achieved. The production data were resolved well, obtaining the data misfit equal to 52 (expected $\approx N/2 = 63$, where N is the number of measurements, see, e.g., Oliver et al. (2008)).

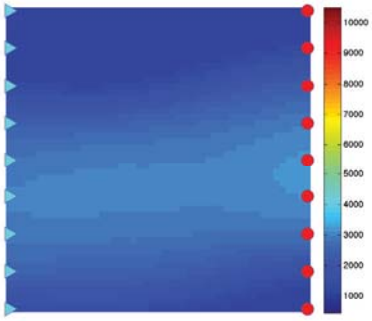


Figure 3 *Initial permeability model.*

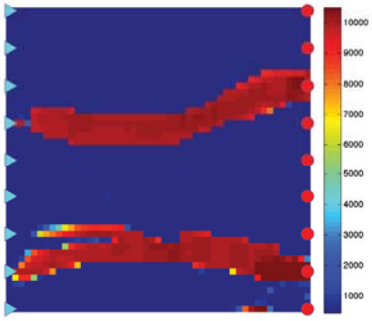


Figure 4 *Solution, 97th iteration.*

Figure 5 demonstrates history matching for injector 4 and producer 3 (wells are numbered starting from the top).

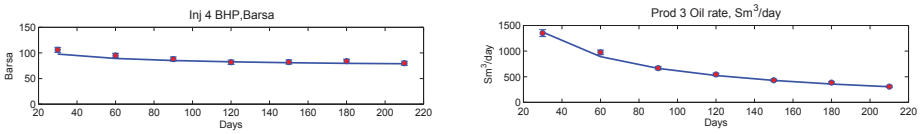


Figure 5 *History matching: observed data (red circles) with error bars and solution response (blue line).*

Conclusions

The proposed approach allows us to solve history matching problem by gradient-based optimization techniques, conserving geological realism of the solution. The differentiable formulation scales down the amount of required forward simulations and can be a valuable approach in modern reservoir management techniques as, for instance, in closed-loop optimization. Besides, the ability to quantify prior probability of history-matched reservoir models allows us to control the quality of reservoir characterization choosing the most reliable solutions.

References

Gao, G. and Reynolds, A.C. [2006] An improved implementation of the LBFGS algorithm for automatic history matching. *SPE Journal,* **11**(1), 5-17.

Guardiano, F. and Srivastava, R. [1993] Multivariate geostatistics: Beyond bivariate moments. *Geostatistics Troia*, **92**(1), 133-144.

Lange, K., Frydendall, J., Cordua, K.S., Hansen, T.M, Melnikova, Y. and Mosegaard, K. [2012] A frequency matching method: solving inverse problems by use of geologically realistic prior information. Accepted to *Mathematical Geosciences.*

Marler, R.T. and Arora, J.S. [2004] Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization,* **26**(6), 369-395.

Oliver, D.S., Reynolds, A.C., Liu, N. [2008] *Petroleum reservoir characterization and history matching*. Cambridge University Press, New York.

Schlumberger GeoQuest [2009] *ECLIPSE reservoir simulator, Technical description*. Houston, TX.

Zhu, C., Byrd, R.H., Lu, P. and Nocedal, J. [1997] L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS),* **23**(4), 550-560.

Appendix A11:
Accounting for imperfect forward models in geophysical inverse problems – exemplified for cross hole tomography

Authors:

Thomas Mejer Hansen, Knud Skou Cordua, Bo Holm Jacobsen, and Klaus Mosegaard

Published in:

Geophysics (Accepted with moderate revisions)

Accounting for imperfect forward models in geophysical inverse problems - exemplified for cross hole tomography.

Thomas Mejer Hansen[*], *Knud Skou Cordua*[*], *Bo Holm Jacobsen*[†] *and Klaus Mosegaard*[*]

ABSTRACT

Inversion of geophysical data rely on knowledge about how to solve the forward problem: That is, computing data from a given set of model parameters. In many applications of inverse problems the solution to the forward problem is assumed to be known perfectly, without uncertainty. In reality though, solving the forward model will almost always be prone to errors, which we refer to as modelization errors. For a specific forward problem, computation of cross hole first arrival travel times, we demonstrate how the modelization error, given a number of different approximative forward models, can be more than an order of a magnitude larger than the measurement uncertainty. We also show that modelization error is strongly linked to the spatial variability of the assumed velocity field, i.e. to the a priori velocity model. We propose a method for generating a sample of the modelization error due to using a simple and approximative forward model, as opposed to a more complex and more correct forward model. Then a probabilistic model of the modelization error is inferred in form of a correlated Gaussian probability distribution. Key to the method is the ability to generate a number of realizations from the a priori model, from which the modelization error can be quantified. The methodology is demonstrated for two synthetic cross hole tomographic inverse problems. Ignoring the modelization error can lead to severe artifacts, which erroneously appear to be well resolved, in the solution of the inverse problem. Accounting for the modelization error leads to a solution of the inverse problem consistent with the actual true model.

INTRODUCTION

Computation or prediction of the measurements, $\mathbf{d}$, given a set of model parameters $\mathbf{m}$ is referred to as solving the *forward problem*. We will consider problems where the forward problem is defined by a possibly non-linear operator g that relates model parameters, $\mathbf{m}$, to data observations, $\mathbf{d}$, such that solving the forward problem can be given by

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

The associated *inverse problem* deals with the problem of inferring information about model parameters, $\mathbf{m}$, given a set of observed data, $\mathbf{d}$, and the forward model, g. The accuracy of the forward model directly influences the solution to the inverse problem.

In many important cases the forward mapping is available only as a costly numerical simulation, i.e. only a more or less accurate approximation to $g(\mathbf{m})$ is available. Thus, only very rarely do wavefields, electromagnetic fields and heatflow fields in inhomogeneous media have a closed analytical form. Moreover, when forward mappings are computer intensive

it becomes interesting to search for a computational less expensive, but typically also less accurate, forward mapping. This introduces an error in the forward model that will lead to erroneous data calculations. We refer to this type of error as a *modelization error*, following Tarantola (2005), which is sometimes also referred to as the *theoretical error*, Tarantola and Valette (1982a).

A common type of modelization error stems from the fact that geophysical data are the response of a 3D earth. A full 3D solution of the forward problem may be computationally too expensive to apply for some problems. The inverse problem may then be solved based on faster 1D, 2D, or $2\frac{1}{2}$D forward models. For example massive amounts of 3D airborne electromagnetic data have been inverted based on a computationally very efficient 1D forward model, see e.g. Viezzoli et al. (2008). Seismic data can in principle be inverted based on accurate 3D full waveform modeling, Tarantola (1988). However, due to computational demands, inversion of reflection seismic data often rely on faster 1D forward modeling codes, see e.g. Buland and Omre (2003).

Other examples of modelization errors are caused by the choice of how one solves the forward problem. Even when 3D modeling codes are used, one may not use a forward model that describes the complete physical problem. Seismic waves, for example, propagates in a 3D viscoelastic anisotropic medium. While methods exist to simulate waveform propagation in such a 3D viscoelastic anisotropic medium (Saenger and Bohlen, 2004), for some problems it may be computationally too expensive to apply, or too complex to handle. Instead one can resort to solve the elastic or acoustic wave-equation. The forward problem of first arrival travel time computation, used in for example cross hole tomography, can be solved by a wide range linear and non-linear types of forward models (wee e.g. Vidale, 1988; Cerveny and Soares, 1992; Jensen et al., 2000; Spetzler and Snieder, 2004). In any case, one will typically introduce a modelization error due to the choice of forward modeling code.

Yet another type of modelization error is linked to the parameterization of the inverse problem, and description of the physical system in which data are recorded. Christiansen et al. (2011) analyze errors due to imperfect system description related to 1D TEM forward modeling, and find that such errors can lead to an error in the estimate of the subsurface layer resistivities that is an order of a magnitude of the true layer resistivities. Cordua et al., (2008; 2009) demonstrate how borehole cavities can lead to significant modelization errors in cross hole georadar tomography if not accounted for.

As we shall discuss in more detail later, the modelization error may be linked to the complexity of the medium that one is trying to infer information about. If the subsurface consists of mainly horizontally stratified layers, then a specific choice of forward model may lead to a small modelization error (see e.g. Fuchs and Müller (1971)), whereas if the subsurface is very inhomogeneous then the same forward model may lead to a large modelization error.

This study provides some general tools by which the modelization error can be quantified and cast into a consistent formulation as an additive Gaussian observation error. The theory and methodologies that follows apply to any inverse problem. In the remainder of this manuscript we will though, as an example, consider the modelization error related to cross-borehole GPR tomography.

First we demonstrate that the use of an approximate method for solving the forward problem, of computing the travel time delay between a source and a receiver, can lead to

modelization errors an order of a magnitude larger than typical measurement uncertainty. It is also demonstrated that the magnitude of the modelization error is closely linked to the type of subsurface heterogeneity.

Then we propose a method that allow generating a sample, in form of a number of realizations, of the modelization error due to the difference between two considered forward models. A probabilistic model of the modelization error is inferred in form of a correlated Gaussian probability distribution with mean $\mathbf{d}_{\mathrm{T}}$ and covariance $\mathbf{C}_{\mathrm{T}}$, $\mathcal{N}(\mathbf{d}_{\mathrm{T}}, \mathbf{C}_{\mathrm{T}})$. This allows accounting for the modelization error as an additive Gaussian observational error.

Finally we show examples of cross hole tomography, demonstrating the effect of disregarding and accounting for the modelization error respectively.

MODELIZATION ERRORS IN CROSS-BOREHOLE GPR TOMOGRAPHY

Cross-borehole tomography is a method widely used in geophysical prospecting for characterizing small-scale variations of near-surface environments. Often the tomographic images are based on first arrival travel times of seismic or electromagnetic signals that are propagated between the boreholes. In this study we consider an example from ground penetrating radar (GPR) cross-borehole tomography. This method has become popular during the last few decades and has various applications such as: Mapping of tunnels and voids (Moran and Greenfield, 1993), mapping of bedrock fractures and fracture zones (Olsson et al., 1992; Lane et al., 1998), estimation of hydrological parameters and delineation of flow paths in the unsaturated zone (Hubbard et al., 1997; Looms et al., 2008a,b), and delineation of geological structures and lithologies (Fullager et al., 2000; Bellefleur and Chouteau, 2001; Tronicke et al., 2004).

A number of sources of modelization error associated with cross-hole georadar tomography have been discussed. Peterson (2001) lists the errors introduced in ground penetrating radar cross-borehole tomography caused by incorrect station geometry and zero time calibration, geometric spreading, transmitter radiation pattern, transmitter amplitude, and high angle raypaths. Cordua et al. (2009) demonstrate that some of these errors (related to e.g. cavities in the borehole walls) result in correlated data errors. They also demonstrate how a model of Gaussian correlated data errors that accounts for these errors can be empirically setup and used in a least squares based inversion approach. Another type of modelization error is associated to the way the forward problem of computing the travel time between a source and a receiver is solved. We will consider this type of modelization error in more detail.

Forwards models for first arrival travel time computation

The forward problem related to travel time tomography consists of computing the time between emitting a seismic or electromagnetic waveform at a source location, and the first arrival time of the propagating waveform recorded at a receiver. By first arrival we specifically mean the first break arrival time. A large number of methods exist to solve the forward problem. We will here briefly discuss some of the most widely used methods to

solve the forward problem related to travel time computation. We will subsequently refer to a specific choice of forward model using a subscript to g as defined in the following.

We will discuss the forward models defined below, by considering three reference velocity models of size 6x3m, with a source located at S=(0.5m, 1.5m), and a receiver located at R=(5.5m, 1.5m), see Figure 1, top.

Model a) is a constant velocity model with a velocity of 0.14 m/ns, Figure 1a, top. Model b) is a realization of a Gaussian random field with mean velocity 0.14 m/ns, variance of 1.44e-4 $(\text{m/ns})^2$, and an exponential covariance model with a direction of maximum continuity of 10 degrees below horizontal, a maximum correlation length of 20m and a minimum correlation length of 2m, see Figure 1b, top. Model c) is a realization of a random function inferred from a binary channel based training image, generated using single normal equation simulation, Strebelle (2002). The velocity outside the channels is 0.1273 m/ns, and 0.1673 m/ns inside the channel. The mean velocity is 0.14 m/ns, see Figure 1c, top.

The sensitivity kernel corresponding to each of the considered forward models for each of the three considered reference models is visualized below each velocity model in Figure 1, where black indicates positive sensitivity (an increase in velocity will cause sa decrease in the travel time), and red indicates negative sensitivity (a decrease in velocity increases the travel time). The sensitivity kernel is simply the fist order Fréchet derivative of the forward model with respect to a given velocity model.

The high frequency (ray) approximation (g_{SR},g_{BR})

The simplest forward model for computing first arrival travel times relies on a high-frequency approximation of the wave equation, often referred to as the 'ray'-approximation. The travel time is found as the delay caused by the signal traveling along the fastest ray path connecting a source and a receiver. This travel time can be efficiently computed using the eikonal solution to the wave equation (e.g. Vidale, 1988). We will refer to this as the '*bending ray*' forward model, g_{BR}. g_{BR} is non-linear as the ray path depends on the velocity model. A further simplification is to assume that the ray path follows a straight line between the source and receiver. This leads to the linear '*straight ray*' forward model, g_{SR}, which has probably been the most widely used forward model in travel time tomography. g_{SR} is identical to g_{BR} in case the velocity field is homogeneous. The second and third row in Figure 1 shows the sensitivity kernels related to the g_{SR} and g_{BR} forward model for the three reference models.

The finite frequency (fat ray) approximation (g_{FR},g_{BFR})

The ray-approximation has become popular because it is simple and computationally cheap. However, the frequency of a propagating wave is always band limited, which results in scattering effects that are not accounted for by the ray-approximation. Therefore so-called fat rays have been considered. For fat rays the travel time is sensitive not only to the travel time delay along the fastest ray path, but to an area around the ray path. (see e.g. Bursink et al., 2008; Dahlen et al., 2000; Spetzler and Snieder, 2004; Jensen et al., 2000; Marquering et al., 1998). For a single frequency source wavelet the sensitivity kernel consists of alternating regions of positive and negative sensitivity, known as Fresnel zones

(Woodward, 1992). Most of the sensitivity for a band-limited source lies within the first Fresnel zone (Jensen et al., 2000), as higher order Fresnel zone sensitivity tend to cancel out. A number of empirical models have been suggested that describe a sensitivity kernel based on the first Fresnel zone (Cerveny and Soares, 1992; Jensen et al., 2000; Husen and Kissling, 2001). Here we specifically make use of the description proposed by Jensen et al. (2000) to compute a sensitivity kernel reflecting the first Fresnel zone. We will refer to this non-linear kernel, where the sensitivity is dependent on the velocity model, as the *bending Fresnel* forward model, g_{BFR}. We will also consider a related linear kernel, obtained by assuming a constant velocity field, and refer to this as the *straight Fresnel* forward model, g_{FR}. The fourth and fifth row in Figure 1 shows the sensitivity kernels related to the g_{FR} and g_{BFR} forward models for the three reference models.

The Born approximation (g_{Born})

Using the Born approximation (considering 1st order scattering) an exact analytical expression for the sensitivity kernel for a point source can be derived for both seismic (Dahlen et al., 2000; Spetzler and Snieder, 2004; Jensen et al., 2000; Marquering et al., 1999; Liu et al., 2009) and electromagnetic wave propagation (Bursink et al., 2008; Liu et al., 2009). Here we will make explicit use of the formulation of the sensitivity kernels given by Buursink et al. (2008), and refer to it as the '*Born*' forward model, g_{Born}.

The sixth row in Figure 1 shows the sensitivity kernels related to the g_{Born} forward model for the three reference models. Note that, as opposed to the previously considered sensitivity kernels, both regions of positive and negative sensitivity are visible. The Born approach is only strictly valid for a homogeneous velocity model, and while sensitivity kernels in principle can be computed for velocity models with small velocity contrasts, the Born approach will fail for larger velocity contrasts. We therefore only consider the sensitivity kernel associated to a homogeneous velocity model. This type of forward model is then linear, independent of the actual velocity model. Note also that sensitivity kernels associated to the Born approximation typically does not reflect the sensitivity of the first break arrival (as we make use of to determine the travel time in this paper) but the time delay associated with maximum cross-correlation between the observed and simulated wavefield. We still consider the sensitivity kernels based on the Born approximation here, as they have previously been used to invert travel time data recorded as first break arrival, see e.g. Liu et al. (2009) and Buursink et al. (2008).

Waveform modeling and first arrival picking (g_{FW})

Perhaps the most precise, and time consuming, approach to solve the forward problem is to use full waveform modeling, followed by picking the first arrival time. We refer to such a forward model as the *FW* forward model, g_{FW}. Specifically we use 2D finite difference waveform modeling (Ernst et al., 2007) with a Ricker source wavelet with a peak frequency of 100 Mhz. First arrival break times are automatically picked using the method proposed by Molyneux and Schmitt (1999). The first order Fréchet derivative, i.e. the sensitivity kernel, related to the g_{FW} forward model is estimated using the perturbation approach (McGillivray and Oldenburg 1990). Here, each model parameter is perturbed slightly and the resulting residual travel time is compared to the travel time related to the reference

model. From this the local gradient, i.e. the 1st order Fréchet derivative can be computed. The resulting sensitivity kernel obtained using the perturbation approach for the three considered reference models is shown in the bottom row in Figure 1.

Note that the sensitivity kernels associated to the g_{SR}, g_{BR}, g_{FR}, g_{BFR}, and g_{Born} forward models completely defines the forward model with respect to the given reference models, as the full Fréchet derivatives are completely given by the first order Fréchet derivative. This is not the case for the g_{FW} forward model, which is sensitive to higher order scattering. The sensitivity kernels shown in the bottom row in Figure 1 are only the first order Fréchet derivatives, and hence only sensitive to first order scattering.

From Figure 1 it is evident that the sensitivity kernels differ, based on the underlying assumptions. The fat ray (g_{FR}, g_{BFR}) and Born (g_{Born}) forward models only consider 1st order scattering, while in reality, higher order scattering may occur. The high frequency forward models (g_{SR}, g_{BR}) do not consider scattering effects at all. Using any of these considered forward models, will result in a possible different travel time than obtained through the finite-difference calculation. Our goal here is not to discuss the validity of each type of forward model, but simply to describe that using any type of forward model will lead to different modelization errors that will affect the computed travel times.

Modelization errors in first arrival travel time computation

We will now analyze the modelization error caused by using a specific choice of forward model. The effect of different kinds of variability in the model space is investigated based on 26 reference models that are grouped into three types of models and shown in Figures 2a-z:

type A: Increasing Gaussian variance Models 1-10 shown in Figures 2a-j are realizations of the same Gaussian random field defined by an exponential covariance model with horizontal range r_{hor} = 15m, a vertical range r_{ver} = 2m, a mean velocity of 0.14 m/ns, but with an increasing standard deviation of, σ = [0, 0.0040, 0.0057, 0.0069, 0.0080, 0.0089, 0.0098, 0.0106, 0.0113, 0.0120] m/ns. Thus, the variability is increasing, while the spatial correlation lengths are kept constant.

type B: Increasing Gaussian spatial correlation length Models 11-20 shown in Figures 2k-t are realizations of the same Gaussian random field defined by an exponential covariance model with a mean velocity of 0.14 m/ns, a constant standard deviation of σ = 0.0120 m/ns, but with an increasing horizontal range of r_{hor}= [0.01, 1.12, 2.23, 3.34, 4.45, 5.56, 6.67, 7.78, 8.89, 10.0] m, assuming a constant anisotropy of factor of r_{hor}/h_{ver} = 5. Thus, the variance is the same, but the spatial correlation lengths are increasing.

type C: Increasing multiple point variability Models 21-26 in Figures 2u-z are realizations of a multiple-point based statistical model inferred from a binary training image (taken from Strebelle, 2001). The actual realization is generated using single normal equation simulation, Strebelle (2002). The location of the channel structure is the same for all 6 models, but the velocity within the channel is decreasing as $v_{channel}$ = [0.13, 0.12, 0.11, 0.10,

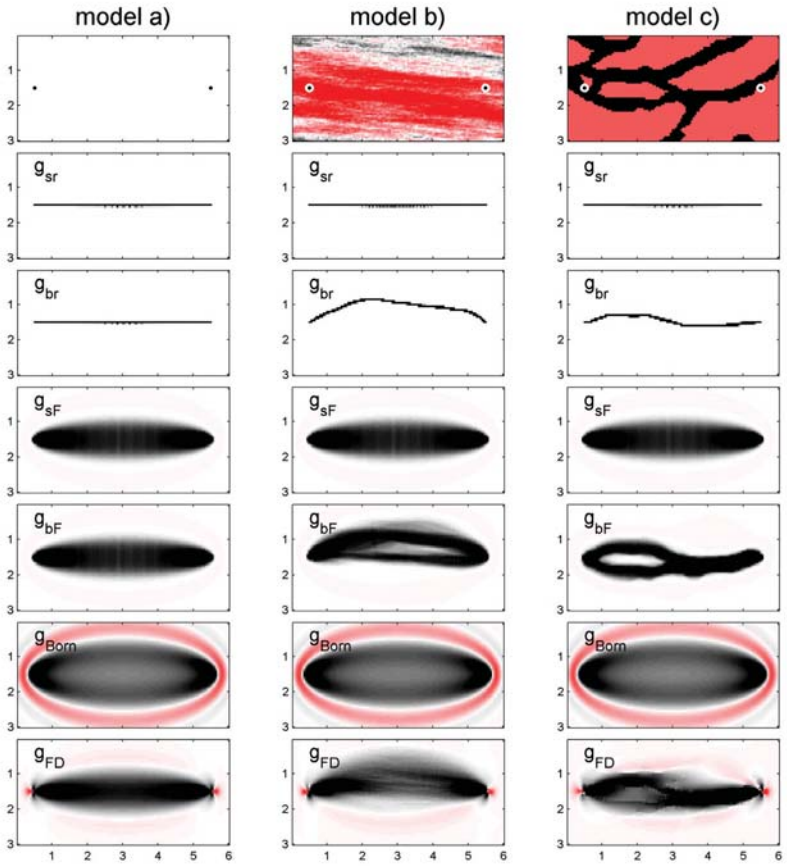


Figure 1: Top row: reference velocity models, a) homogeneous, b) realization of a Gaussian model, c) realization based multiple point statistics inferred from a training image. Red color reflect a high velocity and black a low velocity. Each column show the sensitivity kernels associated to the forward models g_{SR}, g_{BR}, g_{FR}, g_{BFR}, g_{Born}, and g_{FW}. For the sensitivity kernels, black indicates positive sensitivity, and red indicated negative sensitivity.

Accounting for imperfect forward models

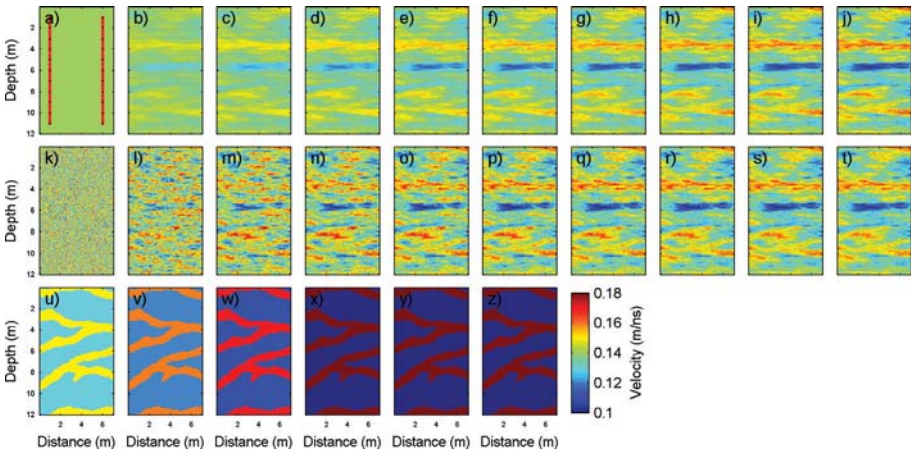


Figure 2: 26 Reference models. a)-j) Models 1-10 of type A (increasing variance). k)-t) Models 11-20 of type B (increasing correlation length and constant variance). u)-z) Models 21-26 of type C (increasing velocity contrast for a binary velocity model). The location of receivers (red dots) and sources (black dots) is denoted on a). See text for more details

0.09, 0.08] m/ns, while the velocity in the region outside the channel increases as $v_{outside}$ = [0.15, 0.16, 0.17, 0.18, 0.19, 0.20] m/ns. Thus the relative velocity contrast at the channel edges is increasing.

For all of these 26 models, a 'reference' travel time data set, $\mathbf{t}_{exact}$, is computed by assuming that the 'exact' forward model, g_{exact} is given by the g_{FW} forward model described above. As a reference geometry we consider the 331 pairs of source and receiver locations shown in Figure 2a. Thus each travel time dataset consists of 331 travel times. In addition, the travel times obtained using the different approximative forward models related to the g_{SR}, g_{BR}, g_{FR}, g_{BFR}, and g_{Born}, will be computed and referred to as $\mathbf{t}_{SR}$, $\mathbf{t}_{BR}$, $\mathbf{t}_{FR}$, $\mathbf{t}_{BFR}$, and $\mathbf{t}_{Born}$, respectively. This will allow quantification of the modelization error as the difference in travel time due to using the approximative forward models and the exact forward model.

Figure 4 shows corresponding modelization error in form of the mean and standard deviation of the error introduced by a specific choice of kernel for the 331 travel times estimates for each of the 26 considered models. The mean can be thought of as a bias, the average mean difference in travel time, and the standard deviation reflect the magnitude of the modelization error.

Figure 4 (top row) illustrates clearly that for a homogeneous model (model 1), all kernels perform equally well. This is no surprise as the kernels are normalized such that the integral of each kernel is the raylength between the source and the receiver. More importantly it clearly shows that as the subsurface variability increases, both the absolute value of the mean and the standard deviation of the modelization error increases using any of the approximations. A similar pattern can be seen for models 21-26, where the velocity contrast

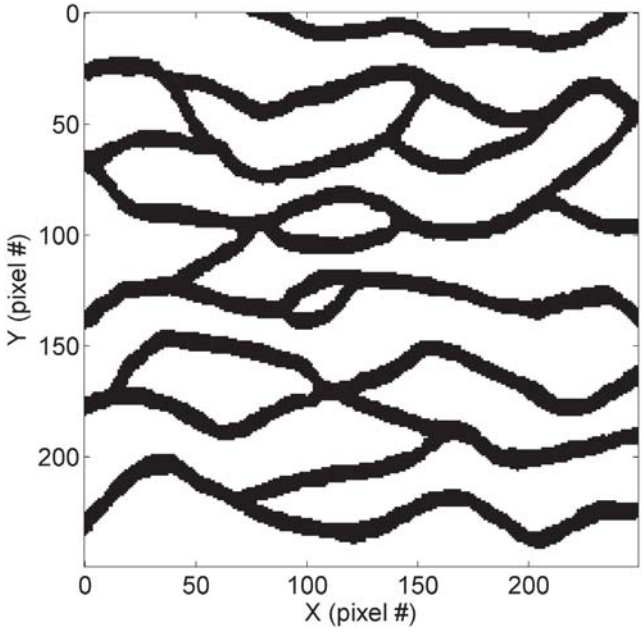


Figure 3: Bimodal training images, used to generate the realizations shown in Figure 2u-z. From Strebelle (2000).

in the binary velocity field is increasing. The absolute value of the mean and the standard deviation is though significantly higher considering the discrete velocity models of models 21-26, than the models based on Gaussian statistics in models 1-10.

Figure 4 (middle row) shows the influence on the bias and variance of the modelization error for increasing spatial correlation lengths. The high frequency approximations (ray sensitivities) are relatively more influenced by the changing correlation length than the ones based on Fresnel volumes. Notice that intermediate spatial wavelengths of the subsurface structures provide larger modelization error than both very high and low correlation lengths. The Fresnel volume sensitivities are less sensitive to the subsurface variability than the ray sensitivities because these sensitivities integrates the velocities of a larger volume of the subsurface.

Using the g_{BR} forward model (high frequency approximation to the wave equation) provides the fastest travel time estimates, since the travel time is computed along the ray that provides the fastest travel time. This is seen as a negative bias in the modelization error. The magnitude of the bias changes from zero (when the velocity model is constant) to about more than -2 ns for the velocity model with channels and maximum velocity contrast, model 26, Figure 4.

Modelization errors related to Gaussian variability of the velocity structures (models 1-20) provides a mean error up to 0.6 ns, and a standard deviation of up to 0.6 m/ns. On the other hand the modelization errors related to variability in the binary velocity structures (models 21-26) provides a mean error up to 12 ns, and a standard deviation of up to 8 m/ns.

Measurement errors in a typical GPR cross-borehole data set is around 0.2 ns to 0.8 ns. Thus for Gaussian type velocity models the modelization error can be up to the order of the uncertainty of the noise model. For the binary velocity distribution the modelization error have the potential to be more than an order of a magnitude larger than the noise model.

Figure 4 thus clearly demonstrates that the modelization error inherent in cross-borehole tomography can be significant, as compared to the measurement error. Furthermore, Figure 4 shows that the magnitude of the modelization error is closely linked to subsurface variability. As the subsurface variability increases so does the magnitude of the modelization error.

In the following we will propose a method to generate a sample of the modelization error, and suggest how to account for such a modelization error when solving inverse problems.

MODELIZATION ERROR AND INVERSE PROBLEM THEORY

The solution to a probabilistically formulated inverse problem is a probability density obtained by combining all available states of information. Prior information on the model and data parameters, obtained independently from physical data, can in general be represented by the a priori probability distribution, $\rho(\mathbf{d}, \mathbf{m})$. Information about the physical relation between data and model parameters can in general be represented by the theoretical probability density, $\Theta(\mathbf{d}, \mathbf{m})$. The combined information, i.e. the solution to the inverse problem, is given by the joint posterior probability defined in the joint data and model space manifold

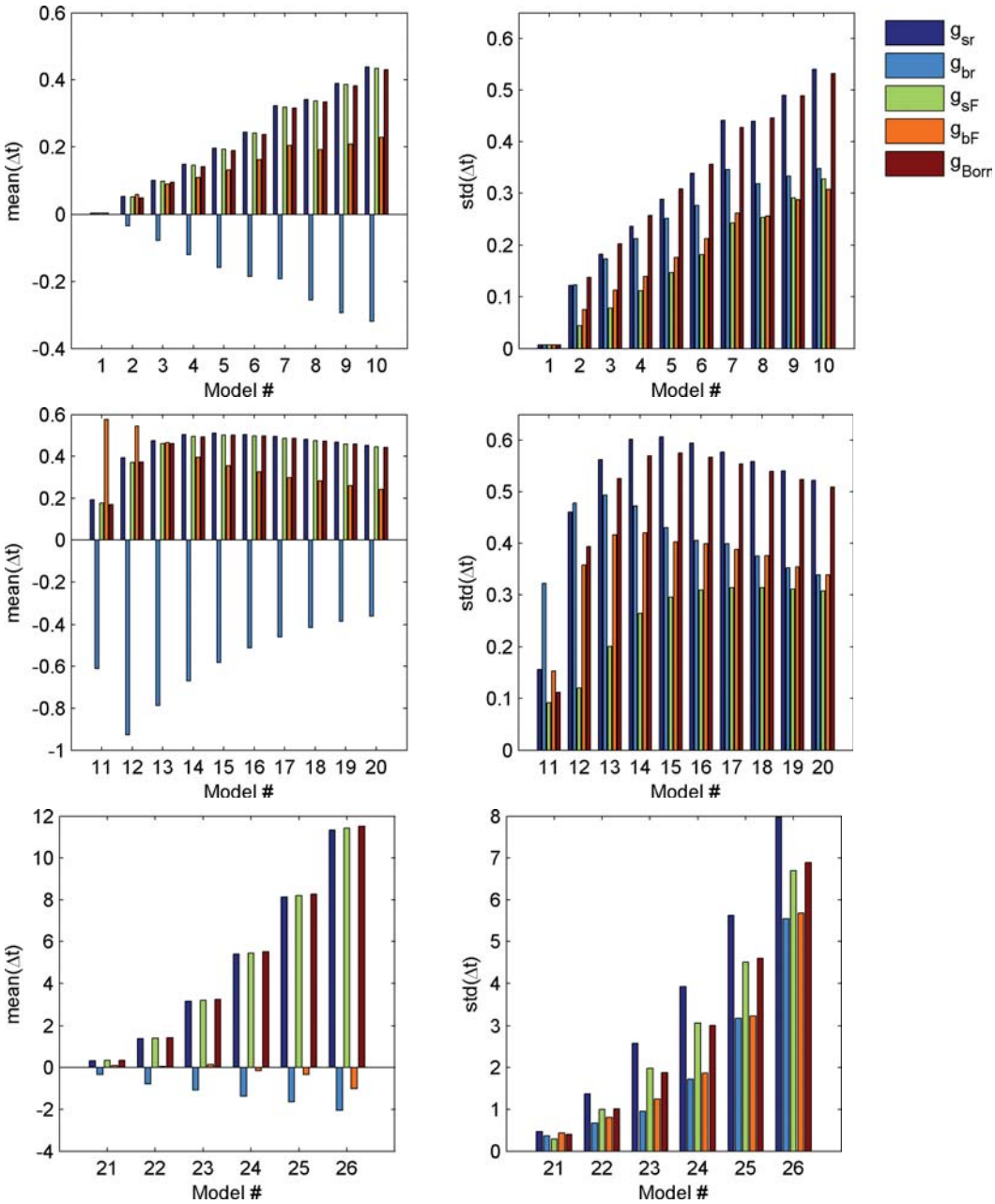


Figure 4: Observed mean (left) and standard deviation (right) of difference in travel times between the travel times obtained using the 'exact' forward model, $\mathbf{t}_{exact}$ and approximate travel time estimates $\mathbf{t}_{SR}$,$\mathbf{t}_{BR}$,$\mathbf{t}_{FR}$,$\mathbf{t}_{BFR}$, and $\mathbf{t}_{Born}$, for velocity models of types A (models 1-10, top row), B (models 11-20, middle row) and C (models 21-26, bottom row).

$\mathcal{D}$ x $\mathcal{M}$:

$$\sigma(\mathbf{d},\mathbf{m}) \;=\; k\;\frac{\rho(\mathbf{d},\mathbf{m})\;\Theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})} \tag{2}$$

The presence of $\mu(\mathbf{d},\mathbf{m})$ in Eqn 2 represents the homogeneous state of information that ensures the parameterization is invariant to changes in coordinate system. Eqn 2 is the most general way to define the solution to an inverse problem in a probabilistic framework, see e.g. Tarantola and Valette (1982) and Mosegaard and Tarantola (2002). One may wish to infer information about the model parameters, $\mathbf{m}$, through the posterior marginal distribution of $\mathbf{m}$ as given by

$$\sigma_{\mathrm{M}}(\mathbf{m}) \;=\; \int_{\mathcal{D}} d\mathbf{d}\;\sigma(\mathbf{d},\mathbf{m}) \tag{3}$$

If it is assumed that $\rho_{\mathrm{D}}(\mathbf{d})$ is obtained independently of $\rho_{\mathrm{M}}(\mathbf{m})$ such that $\rho(\mathbf{d},\mathbf{m}) \;=\; \rho_{\mathrm{D}}(\mathbf{d})\rho_{\mathrm{M}}(\mathbf{m})$, and one assumes the theoretical probability density can be given by

$$\Theta(\mathbf{d},\mathbf{m}) \;=\theta(\mathbf{d}|\mathbf{m})\;\mu_{\mathrm{M}}(\mathbf{m}) \tag{4}$$

where $\mu_{\mathrm{M}}(\mathbf{m})$ is the marginal probability density $\mu_{\mathrm{M}}(\mathbf{m}) \;=\; \int d\mathbf{m}\,\mu(\mathbf{d},\mathbf{m})$, then the solution to the inverse problem of inferring information about the model parameters through the marginal a posteriori probability, $\sigma_{\mathrm{M}}(\mathbf{m})$ can be given by

$$\sigma_{\mathrm{M}}(\mathbf{m}) \;=k\;\rho_{\mathrm{M}}(\mathbf{m})\;L(\mathbf{m}) \tag{5}$$

where the k is a normalization constant and the likelihood function given by

$$L(\mathbf{m}) \;=\; \int_{\mathcal{D}} d\mathbf{d}\frac{\rho_{\mathrm{D}}(g(\mathbf{m}))\;\theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})} \tag{6}$$

$\rho_{\mathrm{D}}(\mathbf{d})$ describes *measurement uncertainties*, typically related to the instrument recording the data. $\theta(\mathbf{d}|\mathbf{m})$ reflects the *modelization uncertainties*, as for example caused by using an imperfect forward model g, as discussed previously.

Figure 5a provides a graphical illustration of the modelization error $\theta(\mathbf{d}|\mathbf{m})$ for a 1 dimensional inverse problem. The red line reflects an assumed error free theoretical relation $\mathbf{d} = g(\mathbf{m})$, and the grayscale density plot reflects $\theta(\mathbf{d}|\mathbf{m})$. Note that in this general case the modelization error is described by a non-Gaussian distribution that is not centered around the theoretical relation $\mathbf{d} = g(\mathbf{m})$. While $\Theta(\mathbf{d},\mathbf{m})$ allows for an arbitrarily complex theoretical probability density, $\theta(\mathbf{d}|\mathbf{m})$ corresponds to quantifying the modelization error by putting vertical uncertainty bars (along with a possible bias) on the theoretical relationship, the red curve (Tarantola, 2005). In this case it is suggested both the bias (the vertical distance from the theoretical relation (red line) to the to the point of maximum likelihood of $\theta(\mathbf{d}|\mathbf{m})$) and the uncertainty increases as $\mathbf{m}$ increases. This resembles, to some degree, the behavior of the modelization error due to the use of imperfect forward models observed in Figure 4.

For the remainder of the text we will consider inverse problems whose solution is given by Eqn. 5, and we will assume that the homogeneous probability density can can be approximated by a constant, such that $\mu_{\mathrm{D}}(\mathbf{d}) \;=\; k$. For more details on the homogeneous probability density function see e.g. Mosegaard and Tarantola (2002) .

Accounting for imperfect forward models

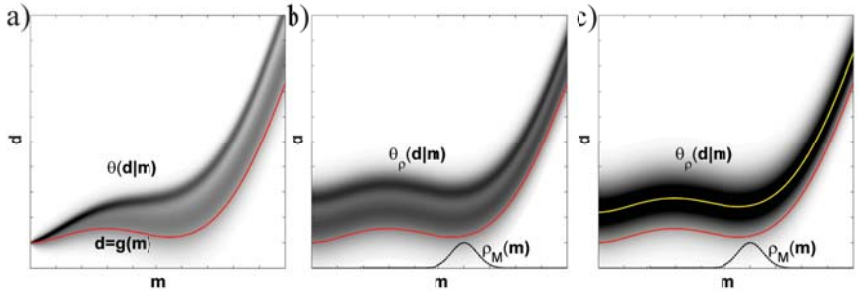


Figure 5: Schematic illustration of the modelization error. Red curve reflect $\mathbf{d} = g(\mathbf{m})$. Grayscale density plot reflect modelization error. a) General description of modelization error, $\theta(\mathbf{d}|\mathbf{m})$. b) Approximate stationary modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$, obtained in the vicinity of $\rho_{\mathrm{M}}(\mathbf{m})$ (dotted line). c) Approximate stationary Gaussian modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$=$\mathcal{N}(\mathbf{d}_{\mathrm{T}}, \mathbf{C}_{\mathrm{T}})$, obtained in the vicinity of $\rho_{\mathrm{M}}(\mathbf{m})$ (dotted line). The yellow curve reflect the maximum likelihood of $\theta_\rho(\mathbf{d}|\mathbf{m})$ for any given $\mathbf{m}$. $\mathbf{d}_{\mathrm{T}}$ is the difference between the red and yellow curve.

Accounting for the modelization error

In general it is not straightforward to evaluate the modelization error, $\theta(\mathbf{d}|\mathbf{m})$, and perform the integration of Eqn. 6.

At one extreme modelization error can be infinitely high, which means that any observed data will lead to the same constant likelihood, $L(\mathbf{m}) = k$. In such a case no information can be inferred from observed data, and the a posteriori probability will simply be identical to the a priori probability, $\sigma_{\mathrm{M}}(\mathbf{m}) = \rho_{\mathrm{M}}(\mathbf{m})$.

Another extreme is to completely ignore the modelization error. When the forward mapping is assumed known precisely, the modelization error can be expressed as $\theta(\mathbf{d}|\mathbf{m}) = \delta(\mathbf{d} - g(\mathbf{m}))$, which reduces Eqn. 6 to

$$L(\mathbf{m}) = \rho_{\mathrm{D}}(g(\mathbf{m})) \tag{7}$$

Evaluation of Eqn. 7 requires the choice of a probability model that describes the measurement uncertainty, i.e. a model that describes how well the forward response of a given model matches the data. If such measurement uncertainty can be described by a Gaussian model, through the covariance $\mathbf{C}_{\mathrm{d}}$, the likelihood can be expressed as

$$L(\mathbf{m}) \;=\; \rho_{\mathrm{D}}(\mathbf{m}) \;=\; k\; exp(-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d}_{obs})^t\; \mathbf{C}_{\mathrm{d}}\; (g(\mathbf{m}) - \mathbf{d}_{obs})) \tag{8}$$

Disregarding the modelization error amounts to assuming a perfectly known noise free relation between $\mathbf{m}$ and $\mathbf{d}$, as indicated by the red line in Figure 5a. If significant modelization errors in reality exist, such as indicated by $\theta(\mathbf{d}|\mathbf{m})$ in Figure 5a, then clearly ignoring such errors may lead to significant artifacts in describing the forward model, and hence, the solution to the inverse problem σ_{M}. This will be further investigated in a case study.

Accounting for Gaussian modelization errors

In the special case where the modelization error can be described by a multidimensional Gaussian probability density, with mean $\mathbf{d}_\mathrm{T}$ and covariance $\mathbf{C}_\mathrm{T}$, $\mathcal{N}(\mathbf{d}_\mathrm{T}, \mathbf{C}_\mathrm{T})$, one can account for both the measurement and modelization error through addition of the covariance models describing measurement error $\mathbf{C}_\mathrm{d}$ and modelization error $\mathbf{C}_\mathrm{T}$ (generalized after Tarantola (1986), page 58):

$$L(\mathbf{m}) = k\ exp(-\frac{1}{2}(g(\mathbf{m}) - (\mathbf{d}_{obs} - \mathbf{d}_\mathrm{T}))^t\ \mathbf{C}_\mathrm{D}\ (\mathbf{g}(\mathbf{m}) - (\mathbf{d_{obs}} - \mathbf{d}_\mathrm{T}))) \tag{9}$$

where $\mathbf{C}_\mathrm{D} = \mathbf{C}_\mathrm{d} + \mathbf{C}_\mathrm{T}$. Eqn. 9 is thus identical to Eqn. 8 except that $\mathbf{C}_\mathrm{d}$ is replaced by the combined covariance model, $\mathbf{C}_\mathrm{D}$, and a bias correction $\mathbf{d}_\mathrm{T}$ is introduced.

This means that one does thus not need to explicitly perform the integration of Eqn. 6. In addition Eqn. 9 is valid for both linear and non-linear inverse problems, as long as the modelization error and the measurement uncertainty can be described by Gaussian statistics, in form of $\mathcal{N}(\mathbf{d}_\mathrm{T}, \mathbf{C}_\mathrm{T})$ and $\mathcal{N}(\mathbf{d}_\mathrm{d}, \mathbf{C}_\mathrm{d})$. If a Gaussian model of modelization error can be established, and is valid, then it can easily be utilized by many types of existing inversion algorithms that can account for Gaussian measurements errors.

In the following we will devise and test methods to approximate the modelization error as an additive correlated Gaussian error. First we suggest methods for generating a sample of the modelization error, that may or may not be Gaussian distributed. Then we suggest how to infer a Gaussian model of modelization error from such a sample of the probability distribution describing the modelization error.

Quantifying modelization error

Generating a sample of the modelization error

Let $\mathbf{M} = [\mathbf{m}'_1, \mathbf{m}'_2, ..., \mathbf{m}'_N]$ represent a sample, in form on N realizations from an a priori probability distribution, $\rho_\mathrm{M}(\mathbf{m})$. Consider two (linear or non-linear) forward models in form of an exact, g_{ex}, and an approximate, g_{app} forward model. Let $\mathbf{M}$ represent N realizations of the prior model $\rho_\mathrm{M}(\mathbf{m})$. The corresponding data, $\mathbf{D}_{ex}$ and $\mathbf{D}_{app}$, for each of the two forward models, can be computed as

$$\begin{aligned} \mathbf{D}_{ex} &= g_{ex}(\mathbf{M}) \\ \mathbf{D}_{app} &= g_{app}(\mathbf{M}) \end{aligned}$$

Each column in $\mathbf{D}_{ex}$ and $\mathbf{D}_{app}$ contains the data, e.g. $\mathbf{d}_{ex,i}$ or $\mathbf{d}_{app,i}$, related to the i'th realization of the a priori probability density, $\rho_\mathrm{M}(\mathbf{m})$. A sample, in form of N realizations, of the (unknown) probabilistic model of the modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$, is now available as

$$\mathbf{D}_\theta = \mathbf{D}_{ex} - \mathbf{D}_{app}. \tag{10}$$

We refer to the obtained sample of the modelization error as a sample from $\theta_\rho(\mathbf{d}|\mathbf{m})$ as it is related to the subset of all models defined by the prior model, $\rho_\mathrm{M}(\mathbf{m})$. To generate a sample of the full modelization error, $\theta(\mathbf{d}|\mathbf{m})$, one would need to consider any model,

Accounting for imperfect forward models

independent of its a priori likelihood. This will in practice not be possible. We therefore choose to assume stationarity of the modelization error, around a priori acceptable models, such that the sample of the modelization error obtained is valid in the vicinity of a priori acceptable models.

Figure 5b illustrates such a modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$, that would be obtained from the full modelization error $\theta(\mathbf{d}|\mathbf{m})$ shown in Figure 5a, using the priori distribution $\rho_{\mathrm{M}}(\mathbf{m})$ shown on the x-axis on Figure 5b. In the vicinity of the prior model, $\theta_\rho(\mathbf{d}|\mathbf{m})$ is a good approximation of the full modelization error $\theta(\mathbf{d}|\mathbf{m})$. For models with small a priori likelihood it is also seen that $\theta_\rho(\mathbf{d}|\mathbf{m})$ does not reflect $\theta(\mathbf{d}|\mathbf{m})$ very well. Note how the the shape of distribution of the modelization error is the same for all values of $\mathbf{m}$, due to the assumption of stationarity. It is only shifted by a constant vertically relative to $\mathbf{d} = g(\mathbf{m})$. For well posed inverse problems, one should not need consider a priori models with very low a priori likelihood, and hence $\theta_\rho(\mathbf{d}|\mathbf{m})$ should be a reasonable approximation $\theta(\mathbf{d}|\mathbf{m})$. If the modelization error is of the stationary nature depicted in Figure 5b, then $\theta_\rho(\mathbf{d}|\mathbf{m})$ will be identical to $\theta(\mathbf{d}|\mathbf{m})$ as $\theta(\mathbf{d}|\mathbf{m})$ is the same for all $\mathbf{m}$.

Approximative estimation of a Gaussian modelization error

In case the sample $\mathbf{D}_\theta$ of modelization errors can be seen as a sample from a Gaussian probability density, we can readily estimate the mean value $\mathbf{d}_{\mathrm{T}app}$ (the bias) and covariance matrix $\mathbf{C}_{\mathrm{T}_{app}}$ of this Gaussian distribution ($\theta_\rho(\mathbf{d}|\mathbf{m}) = \mathcal{N}(\mathbf{d}_{\mathrm{T}app}, \mathbf{C}_{\mathrm{T}_{app}})$) by the following equations

$$\begin{aligned} \mathbf{d}_{\mathrm{T}app} &= [d_{\mathrm{T}app}^1\ ,\ d_{\mathrm{T}app}^2\ ,\ \ldots\ ,\ d_{\mathrm{T}app}^N] \qquad (11)\\ &\quad \text{where} \quad d_{\mathrm{T}app}^j = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}_{ex}^{i,j} - \mathbf{D}_{app}^{i,j}\right)\\ \mathbf{C}_{\mathrm{T}_{app}} &= \frac{1}{N}\,\mathbf{D}_{diff}\mathbf{D}'_{diff} \qquad (12)\\ &\quad \text{where} \quad \mathbf{D}_{diff} = [\mathbf{D}_{ex} - \mathbf{D}_{app} - \mathbf{D}_{\mathrm{T}_{app}}] \end{aligned}$$

and $\mathbf{D}_{\mathrm{T}_{app}} = [\mathbf{d}_{\mathrm{T}app}', \mathbf{d}_{\mathrm{T}app}', \ldots, \mathbf{d}_{\mathrm{T}app}']$.

Figure 5c provides a graphical illustration of such a Gaussian model of modelization error. The gray scale density plot reflect a Gaussian modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$=$\mathcal{N}(\mathbf{d}_{\mathrm{T}_{app}}, \mathbf{C}_{\mathrm{T}_{app}})$, as inferred from inferred from $\mathbf{D}_{ex}$ and $\mathbf{D}_{app}$ using equations 11-12. Again due to the assumption of stationarity, as for Figure 5b, the Gaussian distribution of the modelization error is the same for all values of $\mathbf{m}$, and defined by the variance $\mathbf{C}_{\mathrm{T}}$ and shifted by a constant $\mathbf{d}_{\mathrm{T}}$ relative to $\mathbf{d} = g(\mathbf{m})$, as indicated by the difference between the red and yellow line in Figure 5c.

Whether or not $\mathcal{N}(\mathbf{d}_{\mathrm{T}_{app}}, \mathbf{C}_{\mathrm{T}_{app}})$, is a good description of the actual modelization error $\theta_\rho(\mathbf{d}|\mathbf{m})$ naturally depends on whether the Gaussian assumption is valid or not.

Exact calculation of a Gaussian modelization error in the linear Gaussian case

For linear Gaussian inverse problems, the forward problem can be quantified by the forward operator $\mathbf{G}$, such that $\mathbf{d} = \mathbf{G}\mathbf{m}$. In addition the a priori information on the model param-

eters is described by the Gaussian model $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{M})$. Using $\mathbf{G}$, $\mathbf{m}_0$ and $\mathbf{C}_\mathrm{M}$ one can map the prior information about the model parameters into an a priori Gaussian distribution of data $\mathcal{N}(\mathbf{d}_\rho, \mathbf{C}_\rho)$, with

$$\begin{aligned} \mathbf{d}_\rho &= \mathbf{G}\ \mathbf{m}_0 && (13) \\ \mathbf{C}_\rho &= \mathbf{G}\ \mathbf{C}_\mathrm{M}\ \mathbf{G}' && (14) \end{aligned}$$

$\mathbf{C}_\rho$ should not be confused with data measurement uncertainties, but is simply a description of the covariance between data given the choice of the a priori covariance model and the forward kernel.

Consider an ideal case where an exact expression of the linear forward kernel is given as $\mathbf{G}_{ex}$. Another approximate linear forward kernel is given by $\mathbf{G}_{app}$. The difference in data computed using these two kernels is given by:

$$\begin{aligned} \mathbf{d}_{ex} &= \mathbf{G}_{ex}\ (\mathbf{m} - \mathbf{m}_0)\ +\ \mathbf{G}_{ex}\ \mathbf{m}_0 \\ \mathbf{d}_{app} &= \mathbf{G}_{app}\ (\mathbf{m} - \mathbf{m}_0) +\ \mathbf{G}_{app}\ \mathbf{m}_0 \\ &\Downarrow \\ \mathbf{d}_{ex} - \mathbf{d}_{app}\ -\ [\mathbf{G}_{ex} - \mathbf{G}_{app}]\ \mathbf{m}_0 &= [\mathbf{G}_{ex} - \mathbf{G}_{app}]\ (\mathbf{m} - \mathbf{m}_0) && (15) \\ &\Downarrow \\ \mathbf{d}_{ex} - \mathbf{d}_{app}\ -\ \mathbf{d}_\mathrm{T} &= [\mathbf{G}_{ex} - \mathbf{G}_{app}]\ (\mathbf{m} - \mathbf{m}_0) && (16) \end{aligned}$$

Following Eqn. 16 it is evident that the difference between the exact and approximate estimates of data, $\mathbf{d}_{ex} - \mathbf{d}_{app}$ is linearly related to $\mathbf{m}$ through the difference between the linear kernels, $\mathbf{G}_{ex} - \mathbf{G}_{app}$. Following Eqns. 13-14, a complete description of the Gaussian modelization error $\mathcal{N}(\mathbf{d}_\mathrm{T}, \mathbf{C}_\mathrm{T})$ can now be given as:

$$\begin{aligned} \mathbf{d}_\mathrm{T} &= [\mathbf{G}_{ex} - \mathbf{G}_{app}]\ \mathbf{m}_0 && (17) \\ \mathbf{C}_\mathrm{T} &= [\mathbf{G}_{ex} - \mathbf{G}_{app}]\ \mathbf{C}_\mathrm{M}\ [\mathbf{G}_{ex} - \mathbf{G}_{app}]' && (18) \end{aligned}$$

$\mathcal{N}(\mathbf{d}_\mathrm{T}, \mathbf{C}_\mathrm{T})$ provides an exact description of the modelization error, $\theta_\rho(\mathbf{d}|\mathbf{m})$, caused by using the approximate (and linear) kernel, $\mathbf{G}_{app}$, given a known exact (and linear) kernel, $\mathbf{G}_{ex}$ and the a priori information given by $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{M})$. This is an exact description of the modelization error approximated in Eqns. 11-12.

Example of estimating a Gaussian model of modelization error.

The "exact method" described above applies to linear problems with Gaussian prior and Gaussian data error. The "approximate method" is completely general and provides a sample of the modelization error. The larger the sample, the better statistics about the modelization error can be inferred. We will illustrate this for a linear case where the exact method can serve as the benchmark. We consider two linear forward models defined previously where g_{FR} shall serve as the "true" forward model and g_{SR} serve as the approximative forward model.

Figure 6a-h show the approximate estimates of the covariance of a Gaussian modelization error, $\mathbf{C}_{\mathrm{T}_{app}}$, obtained using Eqn.12 and N=5, 10, 25, 50, 100, 200, 400, and 800 a priori realizations. Figure 6i shows the exact estimate of $\mathbf{C}_\mathrm{T}$ computed using Eqn. 18. The

estimated mean of the Gaussian modelization error, $\mathbf{d}_{\mathrm{T}app}$, obtained using Eqn. 11, is shown in Figure 6j for N =10, 400, and 800 realizations, and compared to the exact estimate of $\mathbf{d}_{\mathrm{T}app}$ computed using Eqn. 18. Figure 6 illustrates how the approximate estimates of $\mathbf{d}_{\mathrm{T}app}$ and $\mathbf{C}_{\mathrm{T}_{app}}$ tend toward the exact estimates of $\mathbf{d}_{\mathrm{T}}$ and $\mathbf{C}_{\mathrm{T}}$ as the number of considered a priori realizations increases.

So in this simple linear case the general nonlinear approach works well and requires relatively few realizations of the modelization error. Below we approach two more realistic problems with non-linear forward models and, for one case, a non-Gaussian prior where our general sampling method can show its potential.

CASES

We have shown, see Figure 4, that the modelization error caused by the use of an imperfect forward model can more than an order of a magnitude larger than the measurement uncertainty. We have also proposed a method to quantify a Gaussian modelization error. Through two case studies we will consider the effect of disregarding and accounting for the modelization error respectively.

In the first case study we will investigate the effect of disregarding the modelization error when solving a non-linear Gaussian inverse problem as a linear inverse Gaussian problem, using least squares type inversion.

In the second case study we explore the use of a fast (and approximate) non-linear forward model to solve a non-linear inverse problem using a non-Gaussian a priori model. In this case a sample of the a posteriori probability density is obtained using the extended Metropolis algorithm.

CASE A : Accounting for modelization errors in linear cross-borehole tomography

Figure 7a shows a reference velocity model which is generated as a realization of Gaussian random field with mean value of 0.14 m/ns, and covariance model with the distance decay known as "spherical" within geostatistics, with a horizontal range of 10 m and a vertical range of 2 meter. From the reference velocity model a reference data set is computed using the forward type model based on finite difference waveform modeling, g_{FW}. The peak frequency of the used Ricker wavelet is 100 Mhz. The same recording geometry is used as in Figure 2a, resulting in 331 computed travel times. In addition three levels of measurement uncertainties, $\mathbf{C}_{\mathrm{d}}$, have been added as uncorrelated Gaussian noise with a standard deviation of 0 (noise free data), 0.1, and 0.4 ns respectively, see Figure 7b. For the remainder of this case study we will consider uncorrelated measurement uncertainty only, and we will refer to a model of uncorrelated measurement uncertainty with a standard deviation of e.g. 0.1 ns as $\mathbf{C}_{\mathrm{d}} = \mathbb{1}\ 0.1^2$. We thus consider three different data sets as observed data, contaminated with different measurement noise.

The 'true' forward' problem is non-linear. We will invert these travel time data using an approximative linear straight ray forward model, g_{SR}, with a known a priori covariance model, $\mathcal{N}(\mathbf{m}, \mathbf{C}_{\mathrm{M}})$, as used to generate the reference model. As we make use of an approx-

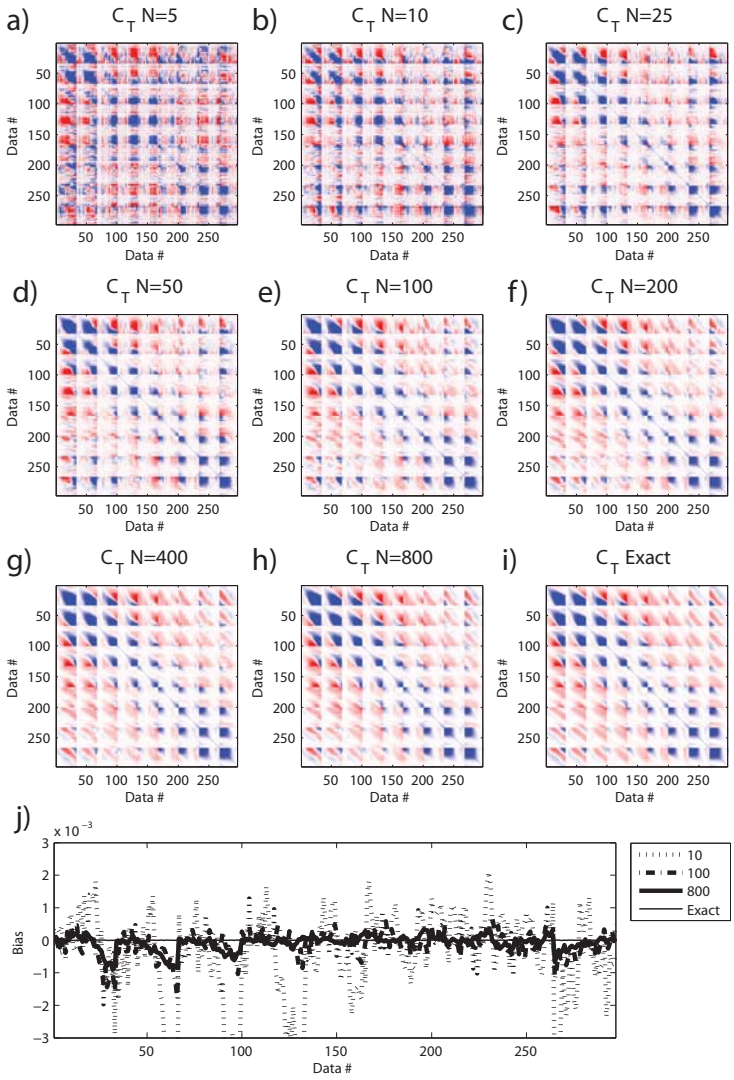


Figure 6: Comparison of the an approximate estimated of $\mathbf{C}_{\mathrm{T}_{app}}$ using Eqn 12 using 1, 135, 361, 475 and 500 realizations from the a priori pdf, to computing the exact $\mathbf{C}_{\mathrm{T}}$ using Eqn 17.

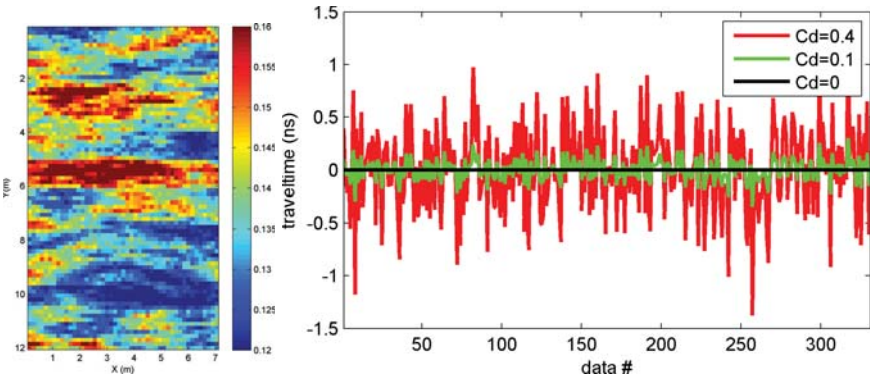


Figure 7: a) Reference velocity model. b) Realizations of the Gaussian models for $\mathbf{C}_{\mathrm{d}}$, with a standard deviation of 0, 0.1 and 0.4 ns.

imation to describe the forward problem a modelization error will be introduced. Treated as a linear Gaussian inverse problem the solution to the inverse problem is described by the Gaussian a posteriori probability density, that can be completely characterized by the a posteriori mean $\widetilde{\mathbf{m}}$ and covariance model $\widetilde{\mathbf{C}}_{\mathrm{M}}$, $\mathcal{N}(\widetilde{\mathbf{m}}, \widetilde{\mathbf{C}}_{\mathrm{M}})$ (see e.g.Tarantola and Valette (1982a)):

$$\widetilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_{\mathrm{M}}\,\mathbf{G}^{\mathbf{t}}\,(\mathbf{G}\;\mathbf{C}_{\mathrm{M}}\;\mathbf{G}'\;+\;\mathbf{C}_{\mathrm{D}})^{-\mathbf{1}}\,(\mathbf{d_0} - \mathbf{G}\,\mathbf{m_0}) \tag{19}$$

$$\widetilde{\mathbf{C}}_{\mathrm{M}} = \mathbf{C}_{\mathrm{M}} - \mathbf{C}_{\mathrm{M}}\,\mathbf{G}^{\mathbf{t}}\,(\mathbf{G}\;\mathbf{C}_{\mathrm{M}}\;\mathbf{G}'\;+\;\mathbf{C}_{\mathrm{D}})^{-\mathbf{1}}\,\mathbf{G}\,\mathbf{C}_{\mathrm{M}} \tag{20}$$

The a posteriori mean model $\widetilde{\mathbf{m}}$ is also the model with maximum a posteriori probability.

Inversion with no modelization error

Initially we will disregard the modelization error ($\mathbf{d}_{\mathrm{T}} = 0$ and $\mathbf{C}_{\mathrm{T}} = 0$) such that $\mathbf{C}_{\mathrm{D}} = \mathbf{C}_{\mathrm{d}}$. That is, we properly account for the three cases of measurement errors, but ignore the modelization error. The linear inverse problem is solved given the three datasets using Eqns. 19-20. Figures 8a,b,c show the corresponding maximum a posteriori model, $\widetilde{\mathbf{m}}$, considering $\mathbf{C}_{\mathrm{d}}$=$\mathbb{1}$ 0, $\mathbb{1}$ 0.1^2, and $\mathbb{1}$ 0.4^2 respectively. Figures 8d,e,f show three realizations from the corresponding a posteriori probability density.

Using noise free data the inversion result is severely affected by disregarding the modelization error, Figure 8a,d. Within the area of high ray coverage (in the middle of the model), both the maximum a posteriori model and realizations from the a posteriori probability density show abnormal velocity variation, both in terms of magnitude and spatial variability. This is a result of fitting the modelization error, which act as noise, and is not accounted for. In order to fit the data according to the assumption of noise free data extreme velocity variations are introduced into the solution space. In addition, comparing the realizations from the a posteriori probability density, many of the structures, which are artifacts and a result of fitting modelization noise, seems to be well resolved. A feature is well resolved if the same feature appears in many realizations of the a posteriori probability

density. Thus ignoring modelization error in this case introduces artifacts, which appear well resolved, into the inversion results.

It is off course quite extreme to invert data without assuming any data noise at all. Considering more realistic measurement uncertainties of $\mathbf{C}_{\mathrm{d}} = \mathbb{1}\ 0.1^2$, the problems referred to above persist. The maximum a posteriori model on Figure 8b shows a low velocity zone at x=3, y=5, surrounded by a high velocity zone. This low velocity zone seems to be well resolved as it can be identified on the 3 realizations of the a posteriori probability density, Figure 8e. However, compared to the reference model, Figure 7a, it is evident, that such a feature does not exist, and is an effect of disregarding the modelization error. As the measurement uncertainty is increased the effect of disregarding the modelization error is reduced, see Figure 8c,f. However, using $\mathbf{C}_{\mathrm{d}} = \mathbb{1}\ 0.4^2$, the low velocity zone still appears as a rather well resolved feature.

Estimating Gaussian modelization error

To estimate the modelization error caused using the by the linear straight ray approximation, g_{SR}, 600 realizations of the a priori Gaussian model, $\mathcal{N}(\mathbf{m}, \mathbf{C}_{\mathrm{M}})$ are generated. For each of the 600 models the forward response is computed using the ideal forward method, g_{FW}, providing $\mathbf{D}_{ex}$. Likewise the approximate linear forward model, g_{SR}, is used to calculate the travel times in $\mathbf{D}_{app}$. Then a Gaussian model of modelization error, $\mathcal{N}(\mathbf{d}_{\mathrm{T}app}, \mathbf{C}_{\mathrm{T}_{app}})$, using Eqns. 17-18. Figure 9a-b shows the estimated covariance $\mathbf{C}_{\mathrm{T}_{app}}$, and mean $\mathbf{d}_{\mathrm{T}app}$, describing the inferred model of Gaussian modelization error.

It has been assumed that the modelization error can be characterized by a Gaussian model, as given by Eqns. 17-18. This may not always be a reasonable assumption. One simple way of checking whether the Gaussian model is reasonable is to simply compare actual realizations of the modelization error, as obtained in Eqn. 10, to realizations of the inferred model of Gaussian modelization error. Figure 9c compares the actual modelization error (black curve) to one realization of the inferred Gaussian modelization error (red) for one realization of the prior model, $\mathcal{N}(\mathbf{m}, \mathbf{C}_{\mathrm{M}})$. A visual comparison suggest that the modeled modelization variability resembles the expected/computed variability, even though the true modelization error seems to have a few outliers that may not be consistent with the Gaussian model.

Inversion with modelization error

We now consider inversion of the travel time data considering both measurement uncertainty, $\mathbf{C}_{\mathrm{d}}$, and at the same time the Gaussian model of modelization error, $\mathcal{N}(\mathbf{d}_{\mathrm{T}app}, \mathbf{C}_{\mathrm{T}_{app}})$, using Eqns. 19 and 20. Inversion results, in form of the maximum a posteriori model, and three realizations from the corresponding a posteriori probability density, can be seen on Figure 10, which can be directly compared to the results of disregarding the modelization error in Figure 8.

From the realizations of the a posteriori probability density, Figure 10d-f, it is evident that spatial variability is similar, using any of the three considered noise models, and similar to the spatial variability of the reference model. The resolution decreases as the measurement uncertainties increase, as should be expected.

Accounting for imperfect forward models

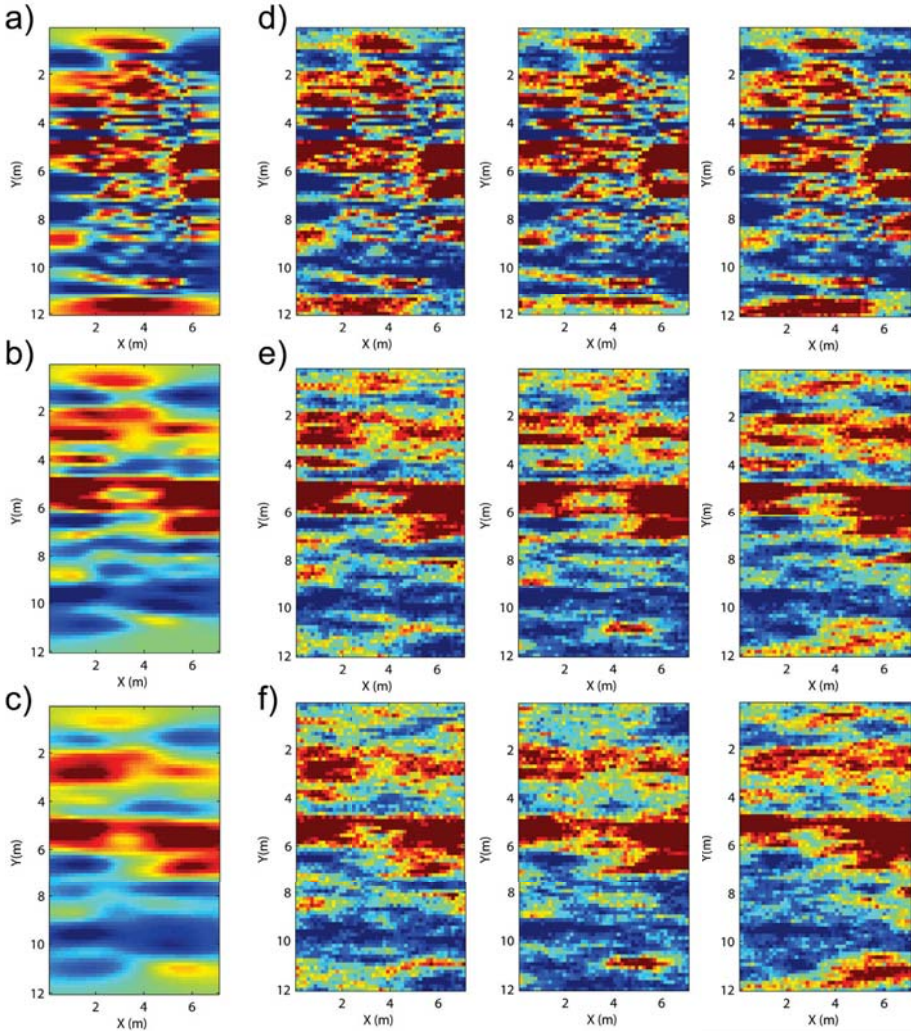


Figure 8: Inversion disregarding modelization error, $\mathbf{C}_{\mathrm{T}} = 0$. a-c) Maximum likelihood model (least squares mean estimate) for $\mathbf{C}_{\mathrm{d}}$=$\mathbb{10}$, $\mathbb{10}.1^2$, and $\mathbb{10}.4^2$. d)-f) 3 realizations from the a posteriori probability density considering $\mathbf{C}_{\mathrm{d}}$=$\mathbb{10}$, $\mathbb{10}.1^2$, and $\mathbb{10}.4^2$.

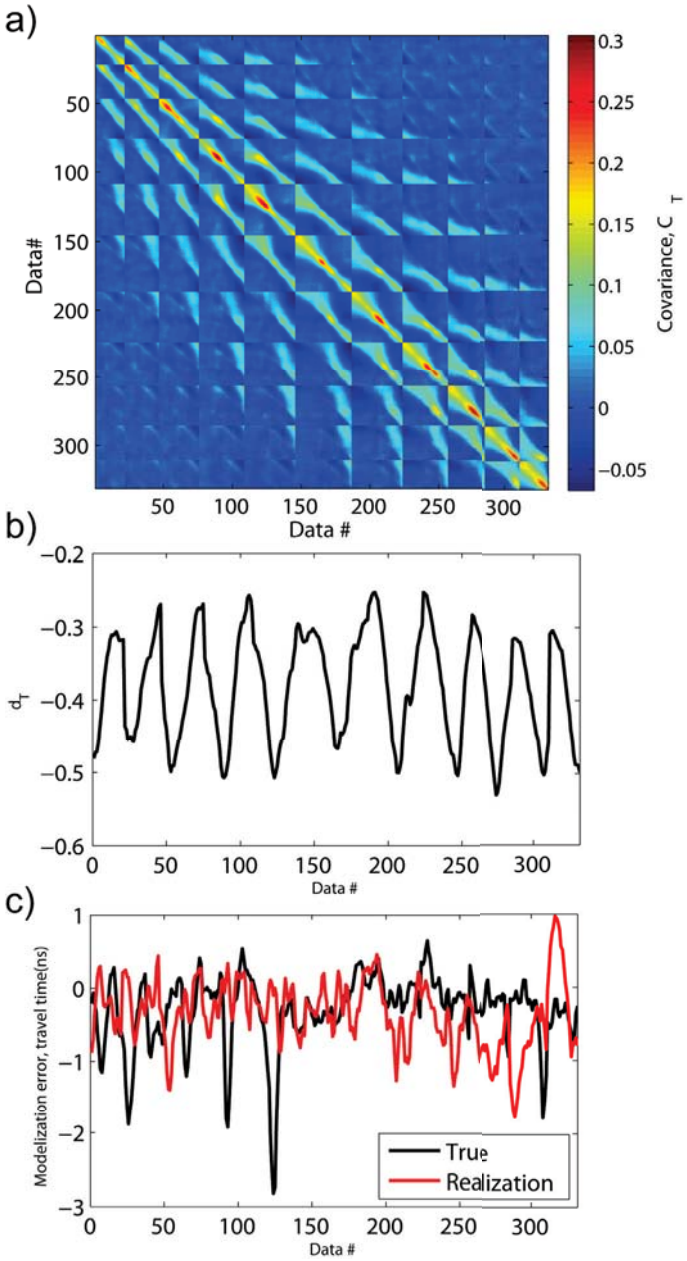


Figure 9: a) Estimated Gaussian model of modelization error, $\mathbf{C}_{\mathrm{T}_{app}}$. b) Estimated mean of modelization error, $\mathbf{d}_{\mathrm{T}_{app}}$. c) Actual modelization error (black) compared to one realization of the inferred modelization error (red) for one realization of $\mathcal{N}(\mathbf{m}, \mathbf{C}_{\mathrm{M}})$.

Accounting for imperfect forward models

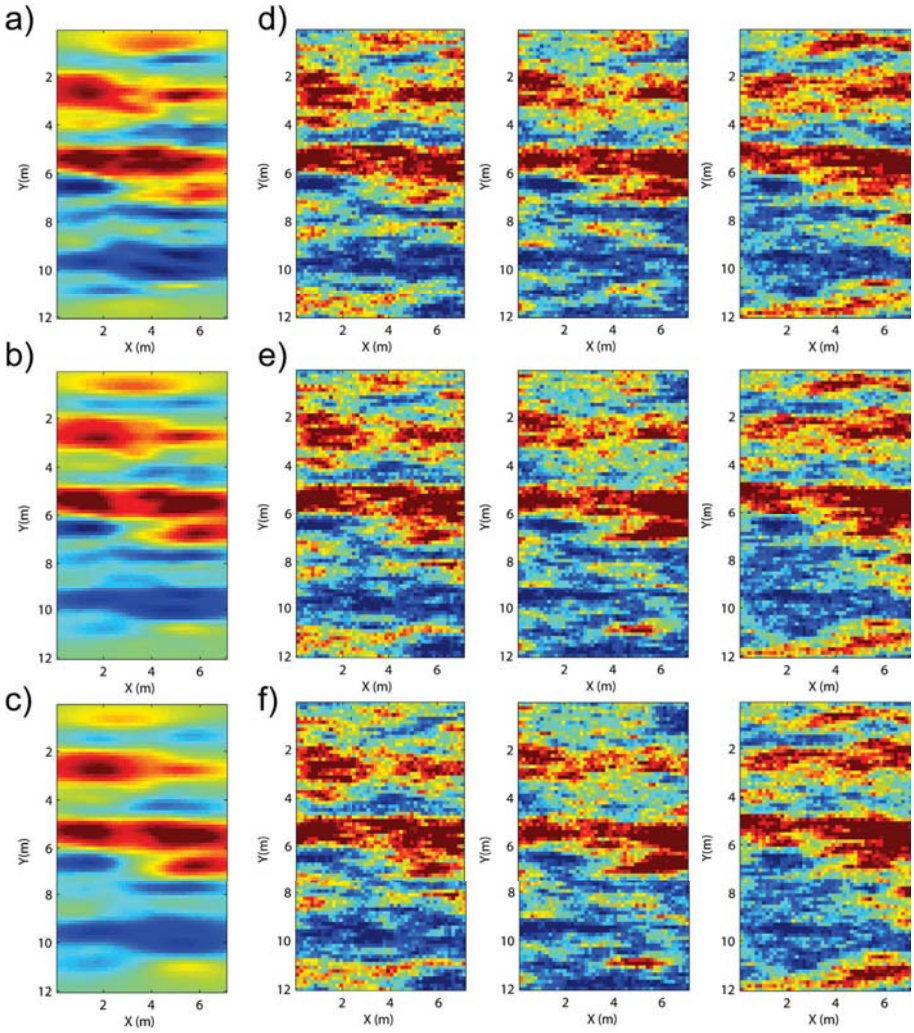


Figure 10: Inversion accounting for estimated modelization error $\mathbf{C}_{\mathrm{T}_{app}}$. a-c) Maximum likelihood model (least squares mean estimate) for $\mathbf{C}_{\mathrm{d}}$=$\mathbb{1}0$, $\mathbb{1}0.1^2$, and $\mathbb{1}0.4^2$. d)-f) 3 realizations from the a posteriori probability density considering $\mathbf{C}_{\mathrm{d}}$=$\mathbb{1}0$, $\mathbb{1}0.1^2$, and $\mathbb{1}0.4^2$. Compare to Figure 8 to evaluate the effect of considering modelization error.

.

It has been shown that it may be possible to solve a nonlinear inverse problem, with a Gaussian a priori model, using a fast approximative linear approach, in a probabilistic sound manner, if the modelization error caused by using an approximative forward model can be described by a Gaussian model. The resulting solution to the inverse problem, in form of the a posteriori probability density, is an approximation to the full inverse problem using the 'true' forward model. This means that model parameters will appear less resolved than if using the exact forward model. However, features that appear well resolved using the approximate inversion approach, should also be well resolved using the full, perhaps computational infeasible, forward model.

Disregarding modelization error can on the other hand produce erroneous inversion results. It can lead to a solution to the inverse problem with extreme model parameter estimates, that appear to be resolved, but which are in reality in conflict with the true subsurface. If modelization uncertainty is ignored one cannot be sure if an apparently well resolved feature is actually resolved, or simply an effect of fitting unknown modelization error as data. It will not be possible to quantify which features are consistent with the true subsurface, without quantifying the modelization error, or solving the inverse problem using the 'exact' forward model.

CASE B : Sampling the a posteriori probability density in case of modelization error.

We will now consider a non-linear inverse problem with a non-Gaussian a priori model. Figure 11 shows a clearly non-Gaussian distributed reference model. It has been generated using the SNESIM algorithm using the image in Figure 3 as training image, Strebelle (2000). The velocity within the thin channel structures is 0.18 m/ns and the velocity of the background material is 0.10 m/ns.

As for case A, a reference data set is computed using the g_{FW} type forward model. A realization of uncorrelated Gaussian noise with a variance of 0.8 ns^2 is added to the reference data to resemble measurement uncertainty, $\mathbf{C}_{\mathrm{d}}$. We now wish to solve the inverse problem of inferring information about the velocity distribution, given 'observed' travel time data, and an a priori model given by the training image in Figure 2d. As a 'non-perfect' forward model we consider the eikonal solution to the wave equation, i.e. the g_{BR} forward model. In this case the forward problem is neither linear nor is the a priori model Gaussian. Therefore we cannot make use of least squares based inversion as in case A. Instead we resort to sampling techniques that, as opposed to providing an analytical description of the a posteriori probability density, will generate a sample of the a posteriori probability density, Mosegaard and Sambridge (2002). We specifically make use of a combination of the extended Metropolis sampler, Mosegaard and Tarantola (1995), and the sequential Gibbs sampler, Hansen et al. (2012), as made available in Hansen et al. (2013a,b).

Quantifying the modelization error

A Gaussian model of modelization error is estimated in a similar manner as was done for case A. Initially 600 realizations is generated from the a priori model, based on the the training image in Figure 3 by the SNESIM algorithm. For each of the 600 models the forward

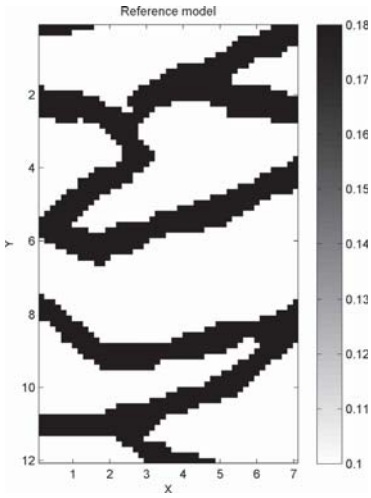


Figure 11: Reference model, case B

response is computed using the ideal forward method, i.e. the g_{FW} type forward model, providing $\mathbf{D}_{ex}$. Likewise the approximate forward model, based on the eikonal solution to the wave equation, g_{BR}, is used to calculate the travel times in $\mathbf{D}_{app}$. Then we estimate a Gaussian model of modelization error, $\mathcal{N}(\mathbf{d}_{\mathrm{T}_{app}}, \mathbf{C}_{\mathrm{T}_{app}})$, using Eqns. 11-12.

Figure 12a shows the estimated bias, $\mathbf{d}_{\mathrm{T}_{app}}$, and figure 12b the estimated covariance, $\mathbf{C}_{\mathrm{T}_{app}}$. Note that the bias itself fluctuates around 1.6 ns, significantly higher than the standard deviation of 0.8 describing the measurement uncertainty.

Sampling the a posteriori probability density

The extended Metropolis algorithm is run for 120000 iterations in two cases: initially assuming no modelization error ($\mathbf{d}_{\mathrm{T}}$=0,$\mathbf{C}_{\mathrm{T}} = 0$), and subsequently we account for the modelization error making use of the estimated Gaussian model of modelization error ($\mathbf{d}_{\mathrm{T}} = \mathbf{d}_{\mathrm{T}_{app}}, \mathbf{C}_{\mathrm{T}} = \mathbf{C}_{\mathrm{T}_{app}}$). The a priori model is the same in both cases.

Figure 13a shows a sample of the a priori probability density in form of 5 different realizations of the a priori distribution. It is apparent that the a priori assumption is that some high velocity channel like structures traverse the model parameter space from left to right, with layers are dipping both up and down from left to right. In addition, some of the channel structures are connected. A corresponding sample from the posterior probability density, in case of disregarding modelization error is shown in Figure 13b, and in case of accounting for the modelization error in Figure 13c.

Disregarding modelization error, Figure 13b, reveals that most realizations within the posterior sample show the same type of features at the same locations. There is relatively

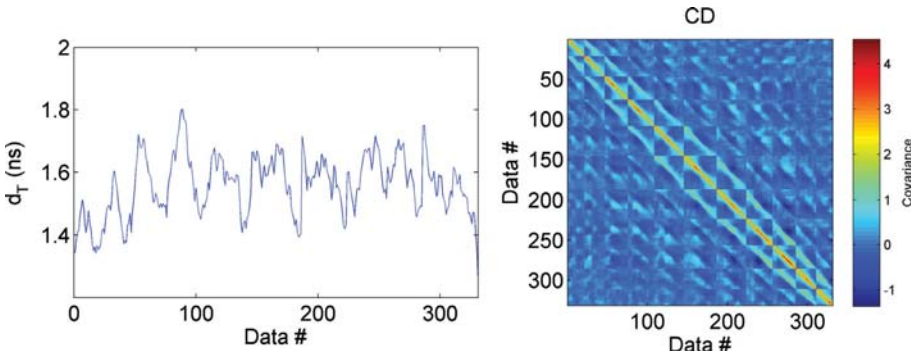


Figure 12: Estimated Gaussian model of modelization error. left) $\mathbf{d}_{\mathrm{T}_{app}}$ right) $\mathbf{C}_{\mathrm{T}_{app}}$

little variability inbetween the realizations. The posterior sample shows considerably more variability between posterior realizations when accounting for the modelization error, Figure 13c. At a first glance it appears as if the posterior sample obtained by disregarding the modelization error provides better resolved features of the model parameter space, than compared to accounting for the modelization error. This is further highlighted in Figure 14, which shows the a posteriori probability of locating a channel for each model parameter. It is apparent that the map of posterior probability of a channel, provides a sharper image disregarding the modelization error, Figure 14a, than when accounting for the modelization error, Figure 14b.

However, comparing Figures 13b and 14 to the reference model, Figure 11, it is clear that when disregarding modelization error, one locates areas with high probability of a channel (and high probability of no channel) that are inconsistent with the reference image. For example, at location (x,y)=(3m,4m) in Figure 14a the probability of a channel is near zero. Yet we know from the reference image that a channel is present at the same location.

On the other hand, when the modelization error is accounted for through the use of the estimated Gaussian model, $\mathcal{N}(\mathbf{d}_{\mathrm{T}_{app}}, \mathbf{C}_{\mathrm{T}_{app}})$, no apparent inconsistencies between the posterior samples and the reference model can be identified. All the features on Figure 14b that appear to well resolved corresponds to real features in the reference image.

Disregarding modelization error can thus introduce features in the a posteriori probability density that appear well resolved, but that are in reality just an effect of mapping the modelization error into the posteriori probability density, as apparently well resolved features. By accounting for the modelization error, the apparent resolution of the posterior sample is reduced, compared to disregarding the modelization error, but features that appear resolved are actually consistent with the subsurface structures.

Note that the MCMC calculations spent 200 times more forward calculations (120000 in all) than the 600 used for the calibration of the modelization error covariance and bias model. One forward computation takes around 67s using g_{FW}, and 0.05s using g_{BR}. The total actual computation time for the setup considered here, using the g_{BR} forward model,

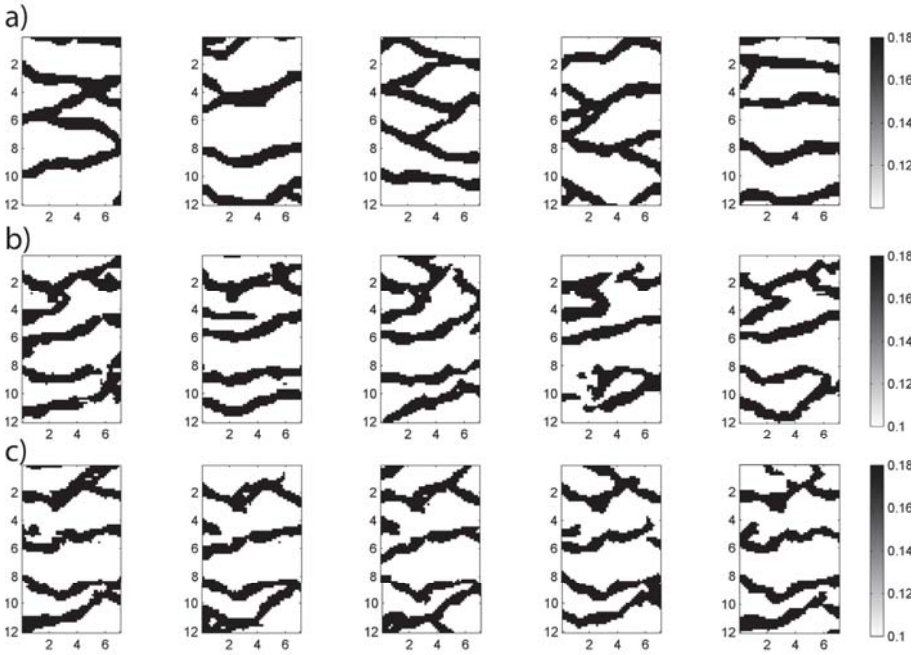


Figure 13: 5 independent realization from the a) prior, b) posterior disregarding modelization error, c) posterior accounting for modelization error.

is around 13 hours (around 11 hours for sampling of the modelization error, and 2 hours for sampling the solution to the inverse problem). The same 120000 iterations would take around 3 months to complete if g_{FW} was used as the forward model, indicating a computational speedup of around 166 using g_{BR} along with a modelization error, compared to using g_{FW} directly. Thus, the use of fast approximative forward models, while at the same time quantifying the associated modelization error, has the potential to allow the use of sampling algorithms with dramatically decreased computational demands.

This second case study illustrates that quantifying the modelization error using a Gaussian model enable sampling of the a posteriori probability density for a non-linear non-Gaussian inverse problem using an approximative, and fast, forward model in a manner consistent with errors introduced by using the approximate forward model. It also demonstrated that if the modelization error is not accounted for, it may introduce features into the posterior probability density that appear well resolved, but are in fact a result of mapping modelization errors into the posterior probability density.

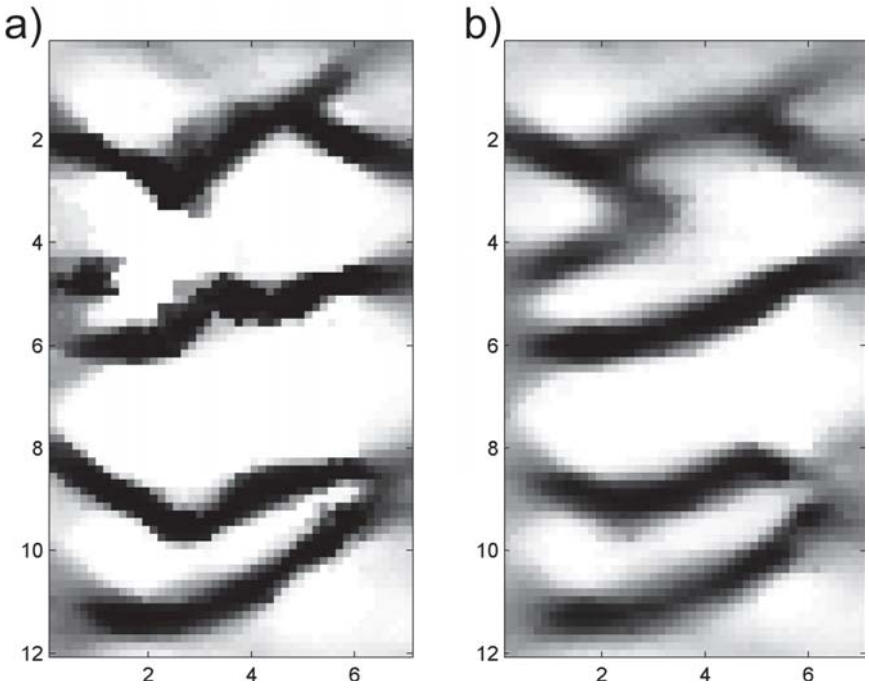


Figure 14: A posteriori probability of a channel disregarding modelization error (a) and accounting for modelization error (b). Black color indicate a probability of 1, and white color indicate a probability of 0.

DISCUSSION

Data errors and modelization errors

Various sources of data errors exist in tomographic travel time data, due to incorrect station geometry, zero time calibration, geometric spreading, transmitter radiation pattern, transmitter amplitude, high angle raypaths, and cavities and small-scale irregularities near the borehole walls (Peterson, 2001; Cordua et al., 2008; 2009). Peterson (2001) argue that such errors in data should be accounted for by pre-processing data in order to remove the data errors prior to inversion. The approach presented here is fundamentally different. We suggest to use the data as they are recorded, and then account for the data errors in form of a probabilistic model of modelization error.

Cordua et al. (2008; 2009) used a probabilistic approach to account for modelization errors as Gaussian static like data errors, related unknown cavities and small-scale clay-enriched zones close to the borehole. They estimate an empirical data covariance model, i.e. a Gaussian statistical model, to quantify such data errors. They show that accounting for such modelization errors remove artefacts from inversion results that stems from subsurface anomalies close to the boreholes. Cordua et al. (2009) calculate the travel time effect of cavities in the borehole walls and small-scale velocity anomalies near the antennae positions as a result of an inadequate forward model that is unable to simulate/model these effects (i.e. modelization errors). They show that these inadequateness lead to correlated data errors that can be accounted for in the inversion through the data covariance matrix.

In this study we propose a general approach to estimate Gaussian modelization error. We have applied the methodology to account for imperfect forward models, but the the method can be generalized for any modelization error for which the cause of the modelization error can be described by a probability distribution, from which realizations can be generated. For example the method can be used to estimate modelization errors stemming from borehole cavities such as investigated by Cordua et al. (2008; 2009). The requirements is simply that one must be able to describe borehole cavities in a statistical model, and one needs to be able to generate realizations from such a statistical model such that the associated data errors can be simulated (by comparing the forward results of models with and without borehole cavities).

In a similar manner uncertainties about the geometry of the sources and receivers can also be considered and quantified as a modelization error, if a statistical model describing such uncertainty can be formulated.

The Gaussian assumption of the modelization error

We specifically make use of a Gaussian probability density function to describe modelization errors. This is convenient as it allows accounting for the modelization error through addition of the covariance models describing measurement uncertainty and modelization error. If another type of parametric probability distribution, as for example a Poisson distribution, better describes the simulated modelization errors $\mathbf{D}_\theta$, such a model should be used instead. The evaluation of the likelihood, equation 6, then becomes more complex.

Applications

The main application of the method we have proposed is that it allows the solution of an inverse problem, using an imperfect forward model, without introducing unwanted artifacts in the solution to the inverse problem, as demonstrated in both case A and B.

If the prior model is Gaussian, and the noise model Gaussian, then one can in principle solve such a non-linear Gaussian inversion problem, using simple linear least squares inversion. One simply needs to use an approximative linear forward model, and compute the corresponding Gaussian model of modelization error. This was shown in case A.

Monte Carlo sampling methods can be computational very expensive, see e.g. Cordua et al. (2012). Therefore one may want to make use of approximative forward models. Without accounting for the associate modelization errors, artifacts may be introduced into computed posterior realizations, and the sampling problem may become hard. The method presented here allow accounting for, at least to some some degree, the modelization error caused by using an approximative forward model, as demonstrated in case B.

Another possible application, not demonstrated here, is to make use of an approximative fast forward model to compute an approximation of the true likelihood function. It has been known for many years that sampling of a probability distribution can be made more efficient if we have an easy-to-sample approximation to the distribution we wish to sample. This was exemplified by Mosegaard and Hansen (2007) who demonstrate the computational efficiency obtained using approximations with a rejection sampler. By quantifying the modelization error as proposed in this paper, any simple forward model may in this way act as an approximation to the full forward problem, and allow for a potential much faster sampling algorithm. It is important to note that using this approach no reduction in resolution of the a posteriori probability density will appear. The a posteriori probability density (related to the hard inverse problem) will be sampled, as well as if using only the 'true' forward model, but due to the use of approximations the computational efficiency will be increased.

CONCLUSIONS

We have proposed a general way to simulate a sample of the modelization error due to the use of an imperfect forward model and/or model description. It can be applied in cases where a statistical model that describes the cause of the modelization error can be quantified. If a Gaussian model can describe the variability of the sample of the modelization error, such a Gaussian model can be straight forwardly inferred from the sample of the modelization error. This will allow accounting for the modelization error by simple addition of the covariance model describing data uncertainty and modelization error.

We have applied the method to inversion of cross hole travel time data, using approximative forward models. First we demonstrated using an example form crosshole georadar, that the modelization error due to the use of approximative forward models can be more than an order of a magnitude larger than the measurement error. We demonstrated the use of approximative forward models to solve a non-linear inverse problem, with a Gaussian a priori model, using classical least squares inversion. We also demonstrated how a non-linear inverse problem, with a non-Gausian a priori model, could be efficiently solved using the extended Metropolis sampler, and an approximative, and fast, forward model. In both cases

we found that disregarding the modelization error led to unwanted artifacts in the generated realizations from the a posteriori probability. If not accounted for, the modelization error will simply be mapped into the a posteriori probability density as unwanted features, that may appear well resolved, but are simply an effect of fitting noise. On the other hand, when accounting for the modelization error, no apparent unwanted features were noticed in the realizations of the a posteriori probability density. In fact, the features that appeared resolved when accounting for the modelization error, was in fact consistent with the features of the known reference model. We also demonstrated that simply increasing the uncorrelated measurement uncertainty in an attempt to account for the modelization error, leads to decreased resolution, while at the same time artifacts appear in realizations of the posteriori probability density.

ACKNOWLEDGMENTS

This work has been partly sponsored by the Danish Research Council for Strategic Research (journal number 10-093934).

REFERENCES

Bellefleur, G., and M. Chouteau, 2001, Massive sulphide delineation using borehole radar: tests at the mcconnell nickel deposit, sudbury, ontario: Journal of applied geophysics, **47**, 45–61.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185.

Buursink, M., T. Johnson, P. Routh, and M. Knoll, 2008, Crosshole radar velocity tomography with finite-frequency Fresnel volume sensitivities: Geophysical Journal International, **172**, 1–17.

Cerveny, V., and J. Soares, 1992, Fresnel volume ray tracing: Geophysics, **57**, 902–915.

Christiansen, A., E. Auken, and A. Viezzoli, 2011, Quantification of modeling errors in airborne tem caused by inaccurate system description: Geophysics, **76**, F43.

Cordua, K., M. Looms, and L. Nielsen, 2008, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data: Vadose Zone Journal, **7**, 263.

Cordua, K., L. Nielsen, M. Looms, T. 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.

Dahlen, F., S. Hung, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes-I. Theory: Geophysical Journal International, **141**, 157–174.

Ernst, J., A. Green, H. Maurer, and K. Holliger, 2007, Application of a new 2d time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, J53.

Fuchs, K., and G. Müller, 1971, Computation of synthetic seismograms with the reflectivity method and comparison with observations: Geophysical Journal of the Royal Astronomical Society, **23**, 417–433.

Fullagar, P., D. Livelybrooks, P. Zhang, A. Calvert, and Y. Wu, 2000, Radio tomography and borehole radar delineation of the mcconnell nickel sulfide deposit, sudbury, ontario, canada: Geophysics, **65**, 1920.

Hansen, T., K. Cordua, M. Looms, and K. Mosegaard, 2013a, SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1, methodology: Computers & Geosciences, **52**, 470–480.

———, 2013b, SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2, Application to cross hole GPR tomography: Computers & Geosciences, **52**, 481–492.

Hansen, T. M., K. C. Cordua, and K. Mosegaard, 2012, Inverse problems with non-trivial priors - efficient solution through sequential Gibbs sampling: Computational Geosciences, **16**, 593–611.

Hubbard, S., J. Peterson Jr, E. Majer, P. Zawislanski, K. Williams, J. Roberts, and F. Wobber, 1997, Estimation of permeable pathways and water content using tomographic radar data: The Leading Edge, **16**, 1623.

Husen, S., and E. Kissling, 2001, Local earthquake tomography between rays and waves: fat ray tomography: Physics of the earth and Planetary Interiors, **123**, 127–147.

Jensen, J., B. Jacobsen, and J. Christensen-Dalsgaard, 2000, Sensitivity kernels for time-distance inversion: Solar Physics, **192**, 231–239.

Lane Jr, J., F. Haeni, and F. Day-Lewis, 1998, Use of time-lapse attenuation-difference radar tomography methods to monitor saline tracer transport in fractured crystalline bedrock: Proceedings of 7th International Conference on GPR, 533–538.

Liu, Y., L. Dong, Y. Wang, J. Zhu, and Z. Ma, 2009, Sensitivity kernels for seismic Fresnel volume tomography: Geophysics, **74**, U35–U46.

Looms, M., A. Binley, K. Jensen, L. Nielsen, and T. Hansen, 2008a, Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data: Vadose Zone Journal, **7**, 238.

Looms, M., K. Jensen, A. Binley, and L. Nielsen, 2008b, Monitoring unsaturated flow and transport using cross-borehole geophysical methods: Vadose Zone Journal, **7**, 227.

Marquering, H., F. Dahlen, and G. Nolet, 1999, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana–doughnut paradox: Geophysical Journal International, **137**, 805–815.

McGillivray, P., and D. Oldenburg, 1990, Methods for calculating Frechet derivatives and sensitivities for the non-linear inverse problem: A comparative study: Geophysical prospecting, **38**, 499–524.

Molyneux, J., and D. Schmitt, 1999, First break timing: Arrival onset times by direct correlation: Geophysics, **64**, 1492–1501.

Moran, M., and R. Greenfield, 1993, Radar signature of a 2.5-d tunnel: Geophysics, **58**, 1573.

Mosegaard, K., and T. Hansen, 2007, Is seismic waveform inversion a hard problem: ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, SPE, 833–842.

Mosegaard, K., and M. Sambridge, 2002, Monte Carlo analysis of inverse problems: Inverse Problems, **18**, 29–54.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447.

———, 2002, Probabilistic approach to inverse problems, *in* International handbook of earthquake and engineering seismology: WHK Lee et al, **81A**, 16, 237–265.

Olsson, O., L. Falk, O. Forslund, L. Lundmark, and E. Sandberg, 1992, Borehole radar applied to the characterization of hydraulically conductive fracture zones in crystalline rock1: Geophysical Prospecting, **40**, 109–142.

Peterson Jr, J., 2001, Pre-inversion corrections and analysis of radar tomographic data: Journal of Environmental and Engineering Geophysics, **6**, 1.

Saenger, E., and T. Bohlen, 2004, Finite-difference modeling of viscoelastic and anisotropic

wave propagation using the rotated staggered grid: Geophysics, **69**, 583.
Spetzler, J., and R. Snieder, 2004, The Fresnel volume and transmitted waves: Geophysics, **69**, 653.
Strebelle, S., 2000, Sequential simulation drawing structures from training images: PhD thesis, Stanford University.
———, 2002, Conditional simulation of complex geological structures using multiple-point statistics: Math. Geol, **34**, 1–20.
Tarantola, A., 1986, Inverse problem theory: Elsevier.
———, 1988, Theoretical background for the inversion of seismic waveforms including elasticity and attenuation: Pure and Applied Geophysics, **128**, 365–399.
Tarantola, A., and B. Valette, 1982a, Generalized nonlinear inverse problems solved using the least squares criterion: Rev. Geophys. Space Phys, **20**, 219–232.
———, 1982b, Inverse problems= quest for information: J. geophys, **50**, 150–170.
Tronicke, J., K. Holliger, W. Barrash, and M. Knoll, 2004, Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation: Water Resources Research, **40**.
Vidale, J., 1988, Finite-difference calculation of travel times: Bulletin of the Seismological Society of America, **78**, 2062–2076.
Viezzoli, A., A. Christiansen, E. Auken, and K. Sørensen, 2008, Quasi-3d modeling of airborne tem data by spatially constrained inversion: Geophysics, **73**, F105.
Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, 15–26.

Appendix A12:

Nonlinear AVO inversion using geostatistical a priori information

Authors:

Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard

Published in:

Proceedings of the 14th Annual Conference of the International Association
for Mathematical Geoscience (IAMG 2010)
Budapest, Hungary
29 August - 2 September 2010

Nonlinear AVO inversion using geostatistical a priori information

Knud Skou Cordua[1], Thomas Mejer Hansen[2], Klaus Mosegaard[3]

[1]Technical University of Denmark. Department of Informatics and Mathematical Modeling
E-mails: [1]knudcordua@gmail.com, [2]tmeha@imm.dtu.dk, [3]kmos@imm.dtu.dk

Abstract

We present a Monte Carlo based strategy for non-linear inversion of seismic amplitude versus offset data. The problem is formulated in a Bayesian framework such that the solution to inverse problem is an a posteriori probability density. A priori information about the problem is defined as a Gaussian probability density. The problem is conditioned by observations of reflected P-waveforms. A nonlinear relation between the model and data based on ray tracing and Zoeppritz equations is considered. As a consequence of these nonlinearities, no closed form expression of the a posteriori probability density can be formulated. Therefore, we apply the Metropolis algorithm in order to obtain an exhaustive characterisation of the a posteriori probability density.

In geophysical inverse problems the model is often discretized into a huge number of model parameters, which results in a high dimensional a posteriori probability density to be sampled. Traditional applications of the Metropolis Algorithm involves that a single model parameter is perturbed according to the a priori information. Each perturbation requests an evaluation of the Metropolis rule, which includes an evaluation of the (often) computationally expensive forward relation. Geophysical model parameters, like subsurface elastic parameters, are often defined in Euclidean space. Therefore, we suggest using geostatistical algorithms, which are designed to handle spatial distributed data set, for defining a priori information about the inverse problems. We demonstrate how the geostatistical Fast Fourier Transform Moving Average algorithm provides a means of multi-parameter perturbations of the a priori information. In this way an efficient sampling using the Metropolis algorithm is obtained. The suggested strategy is tested on a synthetic amplitude versus offset data set.

Keywords: Nonlinear inversion, Metropolis Algorithm, Geostatistics, A priori information, Amplitude Versus Offset

1. INTRODUCTION

Inversion of seismic amplitude versus offset (AVO) data provides information about the subsurface elastic parameters. Reflection seismic prestack recordings contain information about the lithology, which is crucial for the inquiry of hydrocarbon reservoirs that are related to lithological heterogeneities rather than geological structures. Reliable estimates of the elastic parameters are important in order to obtain trustworthy information about the lithology.
In this study we use a Bayesian formulation of the solution to the inverse problem. In this way not only a single estimate but multiple realizations are used to characterize the solution of the inverse problem. This, in turn, provides a means of resolution analysis of the solution (Mosegaard, 1998).

Works of e.g. Buland and Omre (2003) and Kjønsberg et al. (2010) suggest using a Bayesian approach for AVO inversion in order to obtain samples of the a posteriori probability density. In their approaches a weak contract approximation of Zoeppritz equation is used as the forward relation of the problem, which may lead to approximations of the solution in case of high contrast lithological variations.

In this study the forward relation (that relates the elastic parameters to the seismogram) is calculated using ray-tracing and Zoeppritz equation for the P- to P-wave reflectivity coefficients. Reflectivity models obtained in this way are subsequently convolved with the source wavelet in order to obtain the resulting seismogram. Using this forward relation, no approximate pre-inversion angle-dependent data sorting is needed and, instead, common midpoint data can be used directly. Moreover, this fully nonlinear formulation of the problem allows for inversion of the parameters at their actual depth coordinates rather than positions given as two way travel times.

A Monte Carlo based inversion strategy for inversion of AVO data is suggested. This approach combines the Metropolis algorithm with the Fast Fourier Transform Moving Average (FFT-MA) algorithm (Le Ravalec, 2000) and the Gradual Deformation Method (GDM) (Hu, 2000). We suggest using geostatistical algorithms in order to describe and sample a priori information of geophysical inverse problems since these algorithms are designed to model spatial distributed data (Hansen et al., 2008). In this paper we chose to use the FFT-MA algorithm because this algorithm is very efficient at generating Gaussian random fields, even for very large scales. A combination of the FFT-MA algorithm with the GDM algorithm seems to provide a flexible tool for controlling the perturbation step sizes when applied as a priori sampler in conjunction with the Metropolis algorithm (Mosegaard and Tarantola, 1995).

The suggested algorithm is used to sample the a posteriori probability of a 1D elastic model conditioned by synthetic AVO data.

2. METHODOLOGY

The subsurface is discretized into an N layer model. Each layer is represented by a P-wave velocity ($\mathbf{v}_p$), an S-wave velocity ($\mathbf{v}_s$), and a mass density ($\boldsymbol{\rho}$). Thus, the subsurface can be represented by 3xN model parameters, $\mathbf{m}$ (i.e. the model). Data are given by a CMP gather $\mathbf{s}$ of recorded P-wave seismograms. The forward problem that

relates a model with the CMP gather can be represented by the following three equations (Sheriff and Geldart, 1982; Yilmaz, 1987):

$$\boldsymbol{\alpha}(z,t) = g_{raytracing}\left(\mathbf{v}_p(z)\right), \tag{1}$$

$$\mathbf{r}_{pp}(z) = g_{zoeppritz}\left(\boldsymbol{\alpha}(\mathbf{v}_p(z)), \mathbf{v}_p(z), \mathbf{v}_s(z), \boldsymbol{\rho}(z)\right), \tag{2}$$

$$\mathbf{s}(t) = \mathbf{W}\mathbf{r}_{pp}(t), \tag{3}$$

where eq. (1) is solved through a ray-tracing algorithm in the P-wave velocity-depth model in order to determine the two way travel times t and reflection angles $\boldsymbol{\alpha}$ at the N-1 interfaces in the N layered model. The associated P- to P-wave reflection coefficients $\mathbf{r}_{pp}$ at the individual interfaces are obtained through Zoeppritz equations (eq. (2)). The final CMP gather is calculated using the convolution model which relates the source wavelet, $\mathbf{W}$, and the reflectivity model, $\mathbf{r}_{pp}(t)$, with the seismogram, $\mathbf{s}(t)$. Notice that the reflectivity model can be transformed from a function of depth into a function of time since the ray-tracing provides the two way travel times to the individual layers. In practice this step involves an interpolation to an equidistant temporal sampling interval as the travel times and depths are nonlinearly related through eq. (1). Eq. (2) is also a nonlinear relation whereas eq. (3) is linear. Hence, the resulting problem of inferring the elastic parameters as a function of depth is a nonlinear inverse problem.

In a Bayesian formulation the solution to the inverse problem is given as an a posteriori probability density of the model which can be formulated as (e.g. Tarantola, 2005):

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m})L(\mathbf{d},\mathbf{m}), \tag{4}$$

where k is a normalization constant, $\rho_M(\mathbf{m})$ is the a priori probability density, and $L(\mathbf{d},\mathbf{m})$ is the likelihood function. $\rho_M(\mathbf{m})$ describes the probability that the model satisfies the a priori information. $L(\mathbf{d},\mathbf{m})$ describes how well the modelled data explains the observed data given a data uncertainty. In this study both the a priori probability density and the likelihood function are characterized as Gaussian probability densities:

$$\rho_M(\mathbf{m}) \sim N(\mathbf{m}_0, C_m) \tag{5}$$

$$L(\mathbf{d},\mathbf{m}) \sim N(\mathbf{d}_{obs}, C_d) \tag{6}$$

The a posteriori probability density is a non-Gaussian distribution because the forward relation of the AVO responses is nonlinearly related to the elastic parameters of the subsurface. As a consequence no closed form expression can be obtained for the a posteriori probability density. Therefore, the solution to the inverse problem has to be characterized through sampling the a posteriori probability density. One way of obtaining samples from a high dimensional probability density is by using the Metropolis algorithm. The minimum requirement of the algorithm is; 1) a “black box” algorithm that is able to produce samples of the a priori probability density and, 2) a “black box” algorithm that is

able to evaluate the likelihood function for a given model. The flowchart of the Metropolis algorithm is a follows (Mosegaard and Tarantola, 1995):

1) An a priori sampler proposes a sample, $\mathbf{m}_{propose}$, from the a priori probability density, which is a perturbation of a previously accepted model, $\mathbf{m}_{accept}$.

2) The proposed sample is accepted with the probability (known as the Metropolis rule):

$$P_{accept} = \min\left(1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{accept})}\right) \quad (7)$$

3) If the proposed model is accepted, $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{accept}$. Otherwise the $\mathbf{m}_{propose}$ is rejected and $\mathbf{m}_{accept}$ remains.

4) The procedure is continued until a desirable number of models have been accepted. All models accepted by the Metropolis algorithm (counted with possible repetitions) are samples of the a posteriori probability density.

The Fast Fourier Moving Average (FFT-MA) generator is an efficient way to obtain samples from a Gaussian probability density (Le Ravalec, 2000). In this study the FFT-MA algorithm will serve as the “black box” algorithm that provides samples of a Gaussian a priori probability density. The Moving Average strategy uses, unlike a Cholesky decomposition approach, a covariance function instead of a covariance matrix. This approach is considerably superior since realizations of a very large field can be obtained in very short time. The covariance function can be written as a convolution product of a function *g* and its transpose (e.g. Journel and Huijbregts, 1978):

$$c = g * \breve{g} \quad (8)$$

If *g* can be obtained, a Gaussian random field with mean *m* and covariance *c* is obtained as (e.g. Journel and Huijbregts, 1978):

$$y = m + g * z\,, \quad (9)$$

where *z* is a field of Gaussian white noise. The Gaussian random field is in practice obtained using a Fast Fourier Transform (FFT) algorithm to transform *g* into the Fourier domain.

In order to obtain an efficient Metropolis algorithm it is necessary to be able to control the perturbation steps of the a priori probability density. The Gradual Deformation Method (Hu, 2000) is adequate for controlling the degree of perturbation from one Gaussian realization to another. Consider two independent Gaussian fields z_1 and z_2. The GDM algorithm can be used to obtain a new Gaussian field, z, that is a linear combination between two independent Gaussian fields (Hu, 2000):

$$z = z_1 \cos(\pi p) + z_2 \sin(\pi p)\,, \quad (10)$$

where the parameter p takes on values between 0 and ½. Since the FFT-MA algorithm relies on Gaussian white noise it does not matter whether the entire field or only a subarea of z is perturbed. Eq. (10) can be rephrased as:

$$z_{propose} = z_{current}\cos(\pi p) + z_{new}\sin(\pi p), \qquad (11)$$

where z_{new} is either a completely new realization of Gaussian white noise or a new realization of a subarea of the current field $z_{current}$.

Eq. (11) can then be used to generate perturbations of a Gaussian a priori probability density using the following steps (Le Ravalec, 2000):
1) An initial unconditional sample of the a priori probability density is calculated using eq. (9) (i.e. FFT-MA algorithm) based on an initial distribution of normal derivatives, $z_{current}$.
2) A "subarea" of $z_{current}$ is randomly chosen and a new field of Gaussian white noise is substituted with the values within the subarea resulting in a new distribution of Gaussian white noise denoted z_{new}.
3) A new field of Gaussian white noise, $z_{propose}$, based on $z_{current}$ and z_{new} is calculated using eq. (10) (i.e. the GDM algorithm).
4) A new Gaussian field $y_{propose}$ with covariance *c* and mean *m* based on $z_{propose}$ is calculated using eq. (9).
5) By setting $z_{current}$ equal to $z_{propose}$ and repeating step 2 to 4, a desirable number samples from a Gaussian a prior probability density can be obtained.

The size of the subarea in step (2) and the value of the p-parameter govern the exploratory nature of the Metropolis algorithm. For relatively large subareas the perturbation step becomes relatively large. Likewise, relatively large p-values results in large perturbation steps. The size of the perturbation area and the p-value have to chosen subjectively. For a subarea extension larger than one model parameter we suggest using a p-parameter value of ½ and regulate the size of the subarea in order to control the perturbation step. In case the subarea constitutes only a single model parameter the adjustment of the p-value should be used as the regulating parameter.

One strategy for choosing these parameters is to adoptively adjust the size of the subarea or the p-value during the sampling such that a certain acceptance probability of the Metropolis rule is maintained. Gelman et al. (1996) found that the acceptance rate should be around 23% for high-dimensional distributions. For large acceptance rates the algorithm is exploring the a posteriori probability density too slowly. On the other hand, for smaller acceptance rates too many computationally expensive trials are performed. A constant acceptance rate results in a larger perturbation step in the initial (burn-in) period of the sampling than in the subsequent sampling period. In this way the algorithm becomes efficient at detecting areas in the model space of significant probability in the burn-in phase and then adopts to smaller exploration steps. This approach, however, implicitly assumes a constant global smoothness of the a posteriori probability density,

which may result in unexplored areas of the distribution. On the other hand, however, this approach is expected to be appropriate for moderately nonlinear (i.e. nonGaussian) a posteriori probability densities.

Finally, the likelihood function is defined as:

$$L(s) = k \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \left(s^i - s^i_{obs} \right) / \sigma^2 \right), \qquad (12)$$

Where s^i represents the amplitude of the simulated seismic waveforms obtained using eqs. (1) – (3). s^i_{obs} are the sample points of the observed AVO data. σ is the standard deviation of the expected data uncertainty.

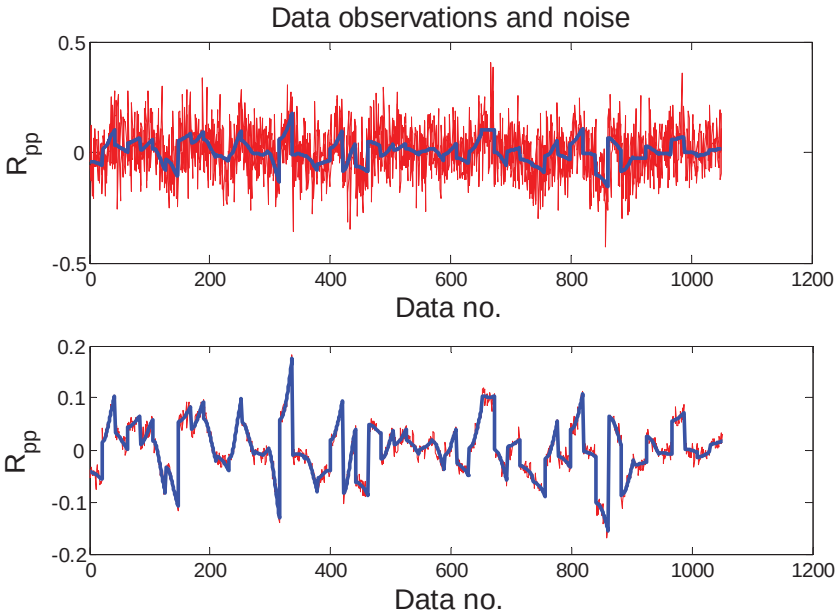


Fig. 1: The synthetic data are plotted as the blue curve. The red curve represents the sum of the synthetic data and the noise component. In the topmost figure a zero mean Gaussian noise component with a standard deviation of 0.1 has been added to the data. In the lower most figure a similar noise component with a standard deviation of 0.01 has been added to data. The resulting signal-to-noise (S/N) ratios of topmost and lowermost signals are 0.24 and 24, respectively.

3. RESULTS AND DISCUSSION

A synthetic reference model is calculated using eq. (9) (i.e. the FFT-MA algorithm). The reference model is a 1000 m deep 50 layer 1D model. The values of the elastic parameters of each layer are seen as the red curve in Figs. 2 and 4. The signals are

recorded with a receiver spacing of 100 m and a maximum offset of 2000 m. Ray-tracing and the associated reflection angels are calculated using the Matlab function *traceray_pp* from the CREWES software package (Margrave, 2003). P- to P-wave reflection coefficients can be calculated using Zoeppritz equations. In a field example the observed data are recorded P-waveforms. Thus, a source pulse has to be convolved with the reflection coefficients in order to obtain the resulting seismogram. In the present synthetic example we will not perform the convolution and the synthetic observed data are given as reflection coefficients. Zero mean Gaussian white noise with a standard deviation of 0.1 and 0.01, respectively, is added to the reflection coefficients (see Fig. 1). A correct assumption about the noise is accounted for in Likelihood function. The results of using the suggested sampling strategy are seen in Figs. 2 – 5. Blue curves show statistically independent realizations of the a posteriori probability density. Green curves are the mean of the (blue) realizations. The dotted black curves in Figs. 2 and 4 show the 0.95 confidence interval of the a priori probability density. Hence, deviations from the a priori distribution can be ascribed to information provided by data.

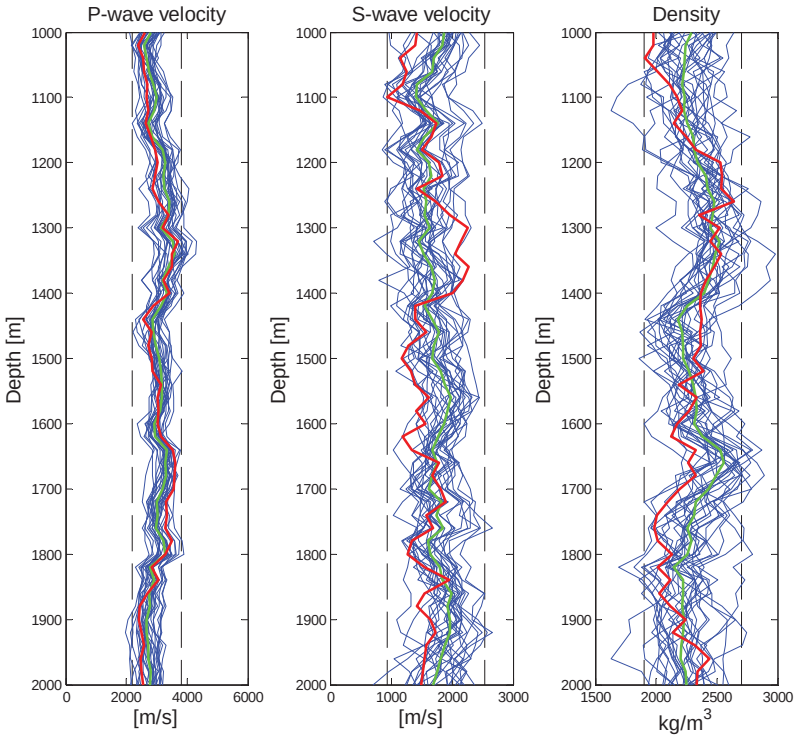


Fig. 2: Results of the inversion using a zero mean normal distributed noise component with a standard deviation of 10^{-1} added to the reflectivity coefficients. The red curve is the "true" reference model, the blue curves are realizations of the a posteriori probability distribution, and the green curve is the mean of the (blue) realizations. The dotted black lines show the 0.95 interval of the a priori probability distribution.

The results show that even though a strong noise component has been added to the data the a posteriori distribution still provides considerable more information about the model parameters than the prior information. In particular, information about the $\mathbf{v}_p$ is resolved. Some information about the density distribution $\boldsymbol{\rho}$ is obtained, whereas least information of the $\mathbf{v}_s$ structure is obtained (see Fig. 2). Fig. 3 shows the result of combining the a posteriori realizations of $\mathbf{v}_p$, $\mathbf{v}_s$, and $\boldsymbol{\rho}$ into $\mathbf{v}_p/\mathbf{v}_s$ ratio, acoustic impedance, and shear impedance. This result reveals that the high spatial frequencies of the $\mathbf{v}_p/\mathbf{v}_s$ ratio are recovered and that the entire spatial frequency spectrum of the acoustic impedance has been resolved. However, a combination of the two least resolved parameters ($\mathbf{v}_s$ and $\boldsymbol{\rho}$) into shear impedance only provides negligible information.

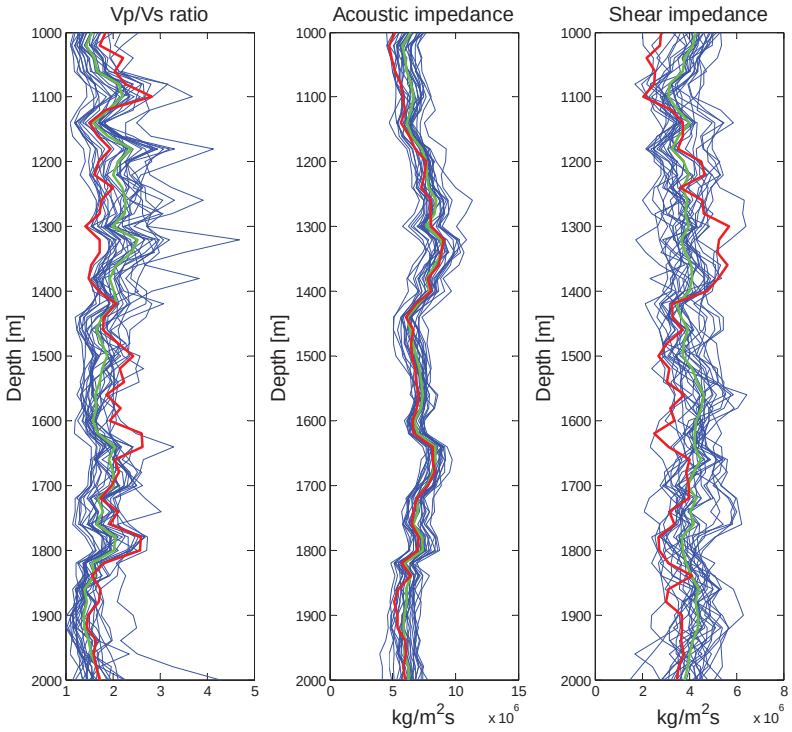


Fig. 3: Results of the inversion using a zero mean normal distributed noise component with a standard deviation of 10^{-1} added to the reflectivity coefficients. The red curve is the "true" reference model, the blue curves are realizations of the a posteriori probability distribution, and the green curve is the mean of the (blue) realizations. The dotted black lines show the 0.95 interval of the a priori probability distribution.

The standard deviation of the noise component is reduced to 0.01. The results of the inversion are depicted in Figs. 4 and 5. The results demonstrate that data are now able to resolve information about both the P- and S-wave velocity as well as the density. However, data seems to provide more information about the high spatial frequencies than compared to the low frequencies. Again it is seen that realizations of the $\mathbf{v}_p/\mathbf{v}_s$ ratio, acoustic impedance, and shear impedance yields an improved resolution as compared to the elastic parameters themselves.

Geostatistical algorithms are often designed for both 2D or 3D application, which is also the case for the FFT-MA algorithm. Hence, a 2D or 3D extension of the suggested inversion strategy can easily be obtained. The computational expensive part of such an extension would be related to a 3D ray tracing.

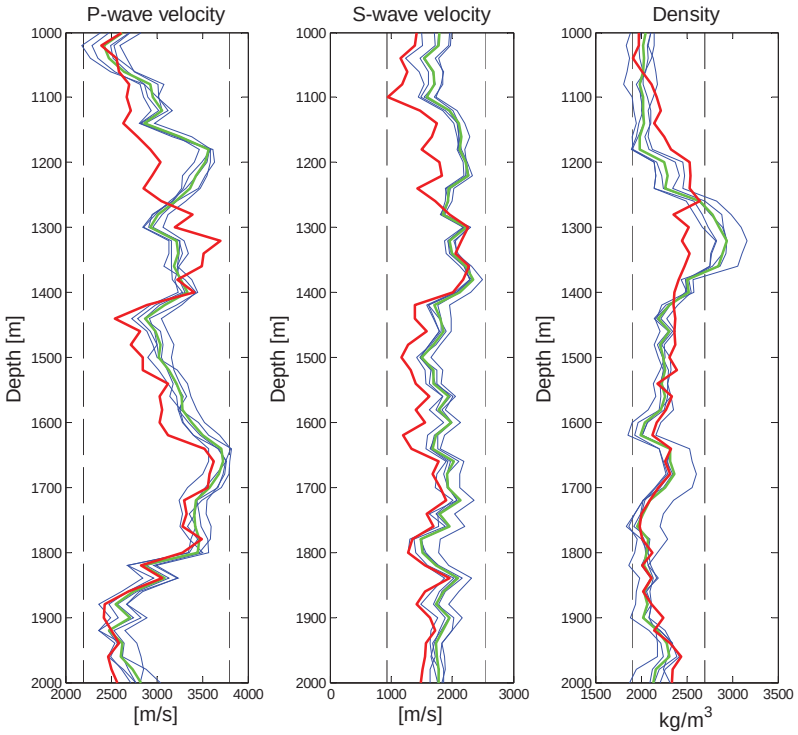


Fig. 4: Results of the inversion using a zero mean normal distributed noise component with a standard deviation of 10^{-2} added to the reflectivity coefficients. The red curve is the "true" reference model, the blue curves are realizations of the a posteriori probability distribution, and the green curve is the mean of the (blue) realizations. The dotted black lines show the 0.95 interval of the a priori probability distribution.

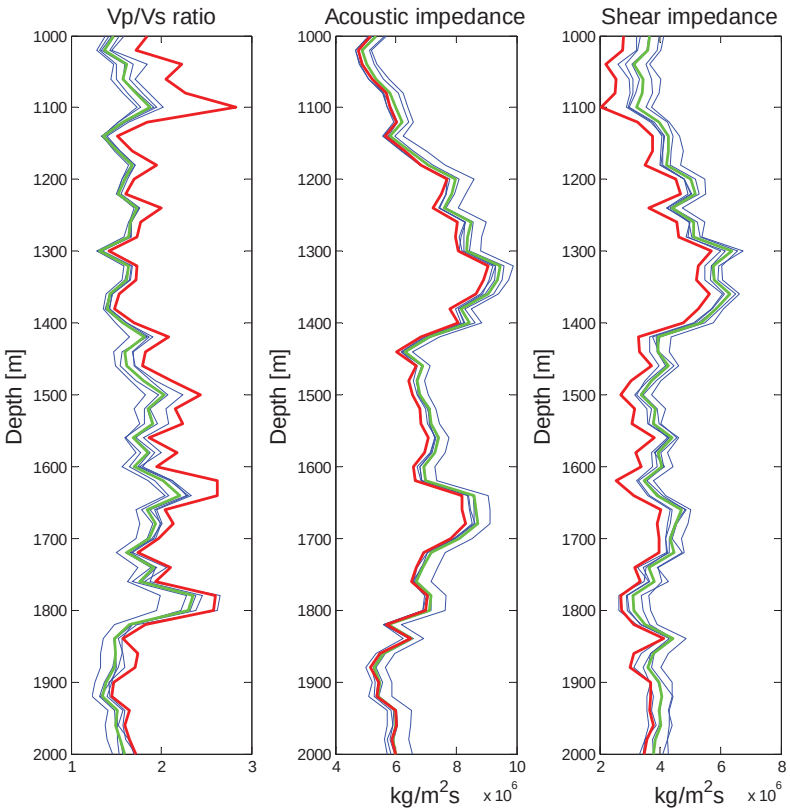


Fig. 5: Results of the inversion using a zero mean normal distributed noise component with a standard deviation of 10^{-2} added to the reflectivity coefficients. The red curve is the "true" reference model, the blue curves are realizations of the a posteriori probability distribution, and the green curve is the mean of the (blue) realizations. The dotted black lines show the 0.95 interval of the a priori probability distribution.

4. CONCLUSION

A nonlinear seismic AVO inverse problem has been formulated using a Bayesian framework. Hence, the solution to the inverse problems is an a posteriori probability density. A strategy combining the FFT-MA algorithm and the Metropolis algorithm has been suggested for sampling of the a posteriori probability density. The strategy has been tested on synthetic data. Even for a S/N ratio of as low as 0.24 the data provide reasonable information about elastic parameters. In particular, a good resolution of the acoustic impedance was obtained. Increasing the S/N ratio to 24 resulted in a considerable improvement in the resolution of all the elastic parameters. Regardless of

the S/N ratio it was observed that the AVO data provided more information about the combined elastic parameters than for the elastic parameters themselves. In particular the results provided a reliable resolution of the acoustic impedance.

5. ACKNOWLEDGEMENT

The present work was sponsored by DONG Energy as part of the research project 'Risk assessment and multiple scenario generation from seismic and geological data'.

6. REFERENCES

Buland, A., H. Omre, 2003, Bayesian linearized AVO inversion. Geophysics, 69(1), 185 – 198.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient Metropolis jumping rules: Bayesian statistics, 5, 599 – 607.

Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2008, Using geostatistics to describe complex a priori information for inverse problems, in J. M. Ortiz and X. Emery, eds., Geostatistics 2008 Chile, vol. 1.

Hu, L. Y., 2000, Gradual deformation and iterative calibration of gaussian-related stochastic models: Mathematical Geology, 32(1), 87–108.

Journel, A.G. and C. J. Huijbregts, 1978, Mining Geostatistics: London, Academic Press. 600 p.

Kjønsberg, H., R. Hauge, O. Kolbjørnsen, A. Buland, 2010, Bayesian Monte Carlo Method for seismic predrill prospect assessment. Geophysics, 75(2), O9 – O19.

Margrave, G.F., 2003, Numerical Methods of exploration seismology with algorithms in Matlab: http://www.crewes.org/ResearchLinks/FreeSoftware/EduSoftware/NumMeth.pdf

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: Inverse Problems, 14, 405 – 426.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of geophysical research, 100, 431 – 447.

Ravalec, M. L., B. Noetinger, L. Y. Hu, 2000, The fft moving average (fft-ma) generator: An efficient numerical method for generating and conditioning gaussian simulations: Mathematical Geology, 32(6), 701–723.

Sherif, R.E., L.P. Geldart, 1984, Exploration Seismology, New York, Cambridge University Press.

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

Yilmaz, Y., 1987. Seismic data processing. Tulsa, Society of exploration geophysicists. 526 p.

Appendix A13:

Multiple Scenario Inversion of Reflection Seismic Prestack Data

Authors:

Thomas Mejer Hansen, Knud Skou Cordua, and Klaus Mosegaard

Published in:

Proceedings of 74th EAGE Conference & Exhibition

Copenhagen, Denmark

4 – 7 June 2012

C006

Multiple Scenario Inversion of Reflection Seismic Prestack Data

T.M. Hansen* (Danish Technical University), K.S. Cordua (Danish Technical University) & K. Mosegaard (Danish Technical University)

SUMMARY

Inversion of prestack seismic data is a highly non-unique inverse problem. Many elastic models exists that fit seismic data observations equally well within the data uncertainty. In a probabilistic formulation of the inverse problem, a full description of such variability is described by the a posteriori probability distribution. Local inversion algorithms quantify the local variability of the a posteriori probability distribution around a 'best' model related to a local (or global) a posteriori likelihood maximum. However, many clusters of models centered around local points of maximum a posteriori likelihood may exist. To locate such local maxima, that may reflect models that fit data within their uncertainty, global inversion algorithms are needed. We present an inversion algorithm, based on Metropolis sampling, that will perform a search in the space of possible model solutions, to identify such different clusters of models. All models will fit data within its uncertainty and obey a possibly complex prior model. We make use of a non-linear forward model, and the inversion does not rely on a known smooth background velocity model.

Introduction

Inversion of prestack seismic data in a probabilistic framework is a challenging inversion problem that has been targeted in the last decades with different approaches. Buland and Omre (1993) solve a prestack inversion problem in a Bayesian context making use of normal moveout corrected prestack seismic data, and a small contrast approximation to Zoeppritz equations, that allow a linear formulation of a convolution problem to describe the relation between elastic parameters and seismic data. Using least squares inversion a Gaussian a posteriori model is computed. This a posteriori model describes the spatial variability around the maximum a posteriori model. Gouveia and Scales (1998) suggest to solve the prestack inversion problem, also in a Bayesian context, using a non-linear optimization method to find the maximum a posteriori model. The uncertainty of the model is then approximated by a local Gaussian a posteriori probability distribution centered at the maximum a posteriori model. Full waveform inversion using the adjoint approach proposed by Tarantola (1982,1984), is one example of an inversion method that allow inference of elastic properties of the subsurface from prestack seismic data. This approach has proven very successful, especially in the case where transmitted waves are recorded. The outcome of this inversion is the maximum a posteriori model that minimized the data misfit. The method does not allow quantification of uncertainty of the solution, nor does the method easily take data uncertainty or existing prior information into account.

The methods described above are local methods, in that one model is found that maximizes a posteriori probability. The uncertainty, if given, is then quantified by a Gaussian probability distribution around the maximum a posteriori model. Thus, if other models exist in the space of acceptable solutions, not on the vicinity of the model with maximum a posteriori probability, then such models will not be part of the posterior uncertainty analysis, and hence this should lead to underestimation of the a posteriori uncertainty.

Here we propose a prestack inversion algorithm that will locate different models that are consistent with data, a noise model and prior information that does not depend on a smooth known background model of the elastic parameters. We will not describe the full posterior probability distribution, but try to locate models centred at point of local maximum a posteriori likelihood, consistent with data, noise and prior information. The development of the algorithm stems from practical observations dealing with manual matching of reservoir properties to seismic reflection data: In order to quantify the commercial viability of a possible reservoir identified from a reflection seismic data set, one wish to identify a number of models whose seismic response match the observed seismic data, within their uncertainties. The method we present is developed to alleviate this process.

Multiple Scenario Sampling

Tarantola and Valette (1982) 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 by combing all the known states of information. In a typical inverse problem the states of information can be described by the a priori and the likelihood pdfs. The 'a priori' pdf, $\rho_M(\mathbf{m})$, describes the data-independent prior knowledge of the model parameters. The likelihood pdf, $L(\mathbf{m})$, 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 a posteriori pdf, $\sigma_M(\mathbf{m})$, which is proportional to the product of the prior and the likelihood:

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

The Metropolis algorithm is a Markov Chain Monte Carlo based algorithm that can be used to sample the a posteriori pdf of non-linear inverse problem in presence of complex a priori information, Mosegaard and Tarantola (1995). This algorithm is at the core of our proposed inversion algorithm

To invert prestack seismic data we propose a sampling strategy where we initially infer the low frequency Vp variations from picked reflection travel times. Then complexity of the considered data is gradually increased, in a workflow as the following:

1. Invert reflection travel time data to obtain smooth Vp velocity trend model. This is done by generating a sample of the a posteriori pdf using the Metropolis algorithm. This produces N independent realizations of the a posteriori pdf for the travel time inversion problem.
2. Invert zero offset data to infer Vp and density. For each of the N low frequent Vp models, a separate Metropolis algorithm is run. This will sample solutions around a local maximum a posteriori model.
3. Invert offset dependent prestack data to infer Vp, Vs and density. Each of the N Metropolis chains is continued allowing variability also of Vs

Synthetic inversion of Pre stack seismic data from the Siri field in the North Sea

Figure 1a show a well log from the North Sea, with a possible reservoir located between two way travel time depths of 1980 ms and 2000 ms (indicated by grey lines in Figure 1a). We will use this as a reference model, from which we compute a synthetic shotgather with offset from 0 m to 3000 m. To compute the prestack shotgather we use a non-linear convolution model, based on ray tracing and the full Zoeppritz equations. We use a Ricker wavelet with maximum peak frequency of 25Hz. For the inference of the low frequent Vp velocity field we have picked the arrival times of 10 coherent reflectors. From these we have estimated a number of RMS velocities associated to a certain time depth.

Initially we consider only travel time data, and assume a priori a relatively long wavelength spatial correlation, in form of a Gaussian covariance model, with range 400 meter and standard deviation of 31 m/s. A Metropolis algorithm is run for 20000 iterations, producing an a posteriori sample of 100 realizations, as shown in Figure 1b. These clearly capture the low frequency variation of the true Vp velocity.

An independent Metropolis algorithm is started in each of 20 randomly selected smooth Vp velocity model. Initially only the zero offset waveform data is considered as data.

A priori models for the spatial variation of Vp and density are inferred from a log located near the reference log, in form of Gaussian based statistics. The prior model for Vp is described by a Spherical covariance model with range 40m and a standard deviation of 200 m/s. The prior model for density is described by a Spherical covariance model with range 40m and a standard deviation of 63 kg/m3. Data noise is defined as uncorrelated Gaussian noise with a standard deviation of 0.03, reflecting a signal to noise ratio of about 30.

In addition a correlation coefficient between Vp and density was found at 0.67. Initially Vp and density is allowed to vary. (Vs is kept constant). After 6000 iterations all 20 running chains found a model that leads to a datafit within the considered data uncertainty. Finally, the full offset dependent seismic shotgather data set is considered as data, and both Vp, Vs and density are inferred. The prior model for the Vs is chosen as a Spherical covariance model with a range of 40 m and a standard deviation of 63 m/s. A correlation coefficient of 0.9 is used to describe the correlation between Vp and Vs. Using the final state of the 20 models from the previous run of the Metropolis algorithm another 20 algorithms is started and run for 6000 realizations.

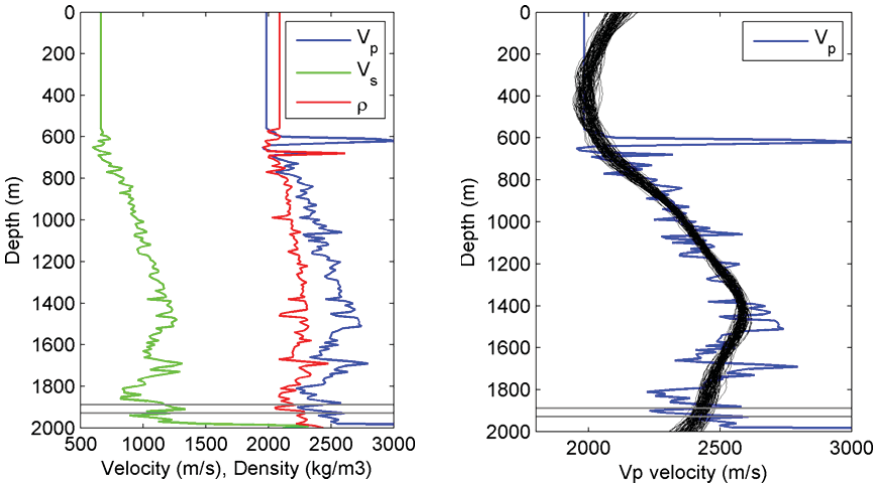


Figure 1 *a,left), Reference Vp ,Vs and density log. b,right), Realizations of the a posteriori pdf, for reflection travel time inversion (black lines) on top of the reference Vp model (blue).*

Figure 2 shows the final 20 realizations in a close up near the reservoir. It can be seen that all 20 models correctly identify the decrease in Vp velocity and Density within the reservoir zone. The Vs velocity however is rather poorly resolved. Figure 3a compare the observed zero offset trace (red) to the zero offset traces computed from 20 accepted models. Figure 3b compare a shotgather computed from one of the accepted models to the reference data set.

Conclusions

The main disadvantage for using a sampling algorithm as proposed is that a large number of forward calculations must be evaluated. This can become computationally unfeasible. However, the computation power keeps increasing, and some of the forward calculations considered here show promise of being very efficient using GPU enabled algorithms.
The are several benefits of using an algorithm such as the proposed. First, the method allows consideration of arbitrarily complex data noise, and very complex a priori information. Second, as the method seeks globally for local maximas it allows inference of a number of classes of models that fit the data and that honor the prior model. Third, as the the inversion relies on a raw (non-nmo corrected data set) it is sensitive to both low frequent and high frequent Vp velocity variations. Finally, the Bayesian formulation of the inversion problem allows linking between elastic properties and rock physics properties such as permeability, porosity and water saturation. In case statistical based rock physics model are available, the presented method should be able to infer for example porosity directly from seismic pre-stack data. Presently we are focusing on incorporating rock physics models to the sampling method, and extending the method to 2D problem.

Acknowledgement

We thank DONG, and specifically Niels Ter-Borch, for financial support, fruitful discussions and permission to publish these data.

References

Buland, A., and H. More, 2003, Bayesian linearized AVO inversion, Geophysics, 68 (1), 185-198.

Gouveia, W. and J.A. Scales, 1997. Resolution of seismic waveform inversion: Bayes versus Occam: Inverse Problems, 13, 323-349.
Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems, Journal of Geophysical Research, 100 (B7), 12431–12447.
Tarantola, A., and B. Valette, 1982, Inverse Problems = Quest for Information: Journal of Geophysics., 50, 159-170.
Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data, Geophysics, 51, 1893-1903.

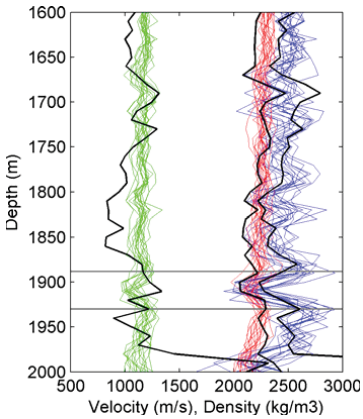


Figure 2 *20 models that result in a datafit that match observed data within its uncertainty.*

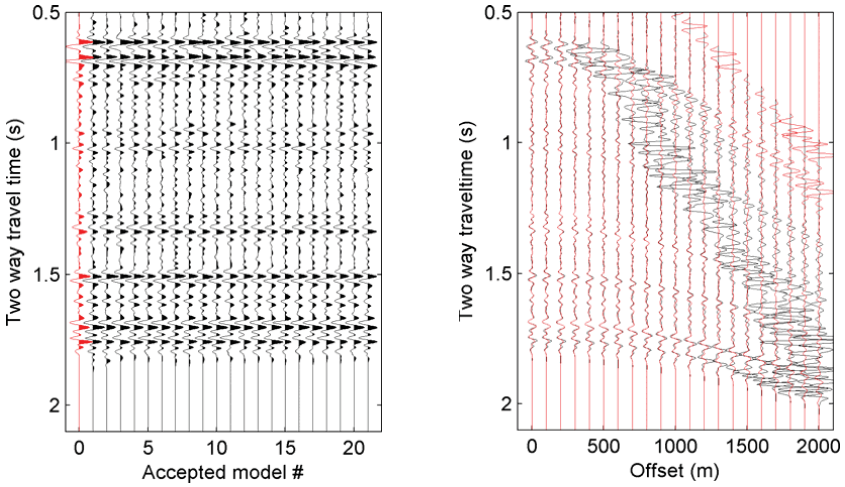


Figure 3 *a, left) Zero offset seismic response from 21 accepted models. The first (red) trace is the zero offset reference trace. b,right) Comparison of the reference shotgather data (black lines) and an (typical) example of the data response from one of the 20 accepted models, of Figure 2*

Appendix A14:

SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 1 - Methodology

Authors:

Thomas Mejer Hansen, Knud Skou Cordua, Majken Caroline Looms, and Klaus Mosegaard

Published in:

Computers & Geosciences

Computers & Geosciences 52 (2013) 470–480

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 1—Methodology

Thomas Mejer Hansen [a,*], Knud Skou Cordua [a], Majken Caroline Looms [b], Klaus Mosegaard [a]

[a] Technical University of Denmark, Center for Energy Resources Engineering, DTU Informatics, Asmussens Alle, Building 305, DK-2800 Lyngby, Denmark
[b] University of Copenhagen, Department of Geography og Geology, Øster Voldgade 10, DK-1350 København K, Denmark

ARTICLE INFO

Article history:
Received 22 June 2012
Received in revised form
7 September 2012
Accepted 10 September 2012
Available online 13 October 2012

Keywords:
Inversion
Nonlinear
Sampling
A priori
A posteriori

ABSTRACT

From a probabilistic point-of-view, the solution to an inverse problem can be seen as a combination of independent states of information quantified by probability density functions. Typically, these states of information are provided by a set of observed data and some a priori information on the solution. The combined states of information (i.e. the solution to the inverse problem) is a probability density function typically referred to as the a posteriori probability density function. We present a generic toolbox for Matlab and Gnu Octave called SIPPI that implements a number of methods for solving such probabilistically formulated inverse problems by sampling the a posteriori probability density function. In order to describe the a priori probability density function, we consider both simple Gaussian models and more complex (and realistic) a priori models based on higher order statistics. These a priori models can be used with both linear and non-linear inverse problems. For linear inverse Gaussian problems we make use of least-squares and kriging-based methods to describe the a posteriori probability density function directly. For general non-linear (i.e. non-Gaussian) inverse problems, we make use of the extended Metropolis algorithm to sample the a posteriori probability density function. Together with the extended Metropolis algorithm, we use sequential Gibbs sampling that allow computationally efficient sampling of complex a priori models. The toolbox can be applied to any inverse problem as long as a way of solving the forward problem is provided. Here we demonstrate the methods and algorithms available in SIPPI. An application of SIPPI, to a tomographic cross borehole inverse problems, is presented in a second part of this paper.

1. Introduction

Inverse problems are abundant in almost any type of scientific research field. An inverse problem occurs when a set of unknown parameters, that describe a physical system, pixel values of an image or some mathematical expression, have to be inferred based on indirect observations of these parameters. Examples of inverse problems are image deblurring, tomographic reconstruction, solutions to certain differential equations, or reconstructing the earth's interior based on surface observations. There are several ways to solve an inverse problem. In a probabilistic formulation the inverse problem can be seen as a way of combining information: Given knowledge about the system (differential equation, physical law, or blurring mechanisms), and a set of observations (signal intensities, pixel values, gravity field), and some prior expectations about the parameters, the goal is to quantify how probable a number of possible scenarios are of explaining the observations and the prior information. A successful probabilistic inversion will, in principle, locate all solutions to the problem and assign a probability to each scenario given the information at hand.

In this paper we present a Matlab[1] toolbox (SIPPI), compatible with Gnu Octave,[2] that can be used to solve inverse problems in a probabilistic formulation. In this formulation the solution to the inverse problem is a probability density function (pdf) referred to as the *a posteriori* pdf that describe all information available about a system. While the toolbox is generally applicable to inverse problems, it has been designed specifically for geophysical inverse problems, where the model parameters typically describe a 1D–3D space, such as for example the subsurface of the earth.

Initially we lay out the theory of probabilistically formulated inverse problems. Then we show how so-called *a priori* information

* Corresponding author. Tel.: +45 45253086; fax: +45 45882673.
E-mail addresses: tmeha@imm.dtu.dk (T. Mejer Hansen), kcor@imm.dtu.dk (K. Skou Cordua), mcl@geol.ku.dk (M. Caroline Looms), kmos@imm.dtu.dk (K. Mosegaard).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.09.004

[1] http://mathworks.com/
[2] http://www.gnu.org/software/octave

about the model parameters, and uncertainty of data observations can be specified. Finally we show how realizations of the a posteriori pdf can be generated using least squares based methods, and sampling techniques such as rejection sampling and Metropolis sampling.

In the second part of this paper we demonstrate the application of SIPPI to a cross borehole traveltime tomographic inverse problem (Hansen et al., this issue).

2. Probabilistic inverse problem theory

Consider some data, $\mathbf{d}$, which are indirect measurements of some model parameters, $\mathbf{m}$, describing a system, such as for example the subsurface of the Earth. Let $\mathbf{d}$ and $\mathbf{m}$ be related through the function g:

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

The above equation, referred to as the forward problem, can be solved with various degrees of accuracy for a number of physical problems.

Inversion of geophysical data amounts to infer information about the model parameters, $\mathbf{m}$, given some data, $\mathbf{d}$, the forward relation between model parameters and data, g, and a priori existing knowledge about the model parameters. Such an inverse problem can be solved in a variety of ways. In this paper we will deal with the general probabilistic formulation of inverse problems. Note that many types of deterministic inversion methods can be formulated as special cases of the probabilistic inverse theory as we consider here.

Tarantola and Valette (1982b) formulate a probabilistic approach for solving inverse problems where all available states of information are described by pdfs. The solution to the inverse problem is the pdf that combines known states of information. In a typical inverse problem the states of information can be described by the *a priori* pdf and the *likelihood* function. The a priori pdf, $\rho M(\mathbf{m})$, describes prior knowledge about the model parameters. The likelihood function, $L(\mathbf{m})$, is a probabilistic measure of how well a given model $\mathbf{m}$ explains the observed data.

The general solution to such a probabilistically formulated inverse problem is the *a posteriori* pdf, which is proportional to the product of the a priori pdf and the likelihood function:

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

where k is a normalization constant and the likelihood is given by

$$L(\mathbf{m}) = \int_{\mathcal{D}} d\mathbf{d}\,\frac{\rho_{\mathrm{D}}(g(\mathbf{m}))\theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})} \tag{3}$$

$\rho_{\mathrm{D}}(\mathbf{d})$ describes *measurement uncertainties*, typically related to uncertainties in the instrument that records the data. $\theta(\mathbf{d}|\mathbf{m})$ describes the *modelization error*, i.e. the error caused by using an imperfect forward model g or an imperfect parameterization. $\mu_{\mathrm{D}}(\mathbf{d})$ describes the homogeneous state of information that ensures that the parameterization is invariant to changes in the coordinate system. For the reminder of the text we shall assume that $\mu_{\mathrm{D}}(\mathbf{d})$ can be approximated by a constant. For more details on the homogeneous pdf, see e.g. Mosegaard and Tarantola (2002).

The a posteriori pdf describes the distribution of models consistent with the combined states of information given by the a priori model and the data.

The probabilistic formulation of inverse problems allows utilization of the movie strategy advocated by Tarantola (2005), who suggest to visualize and compare a sample from the a priori pdf and the a posteriori pdf, respectively, as movies. The 'rior movie' will make it apparent what prior choices have been made. The difference between the prior and the posterior movie will emphasize the effect of using data.

2.1. The linear inverse Gaussian problem

Consider a linear forward problem, where the data $\mathbf{d}$ is linearly related to the model parameters $\mathbf{m}$ using the linear operator $\mathbf{G}$, such that $\mathbf{d} = \mathbf{G}\mathbf{m}$. Let $\mathcal{N}(\mathbf{a},\mathbf{A})$ refer to a Gaussian distribution with mean $\mathbf{a}$ and covariance $\mathbf{A}$. If in addition both the a priori model $\mathcal{N}(\mathbf{m}_0,\mathbf{C}_{\mathrm{M}})$, the noise model $\mathcal{N}(0,\mathbf{C}_d)$ and the modelization error $\mathcal{N}(0,\mathbf{C}_T)$ can be described by a Gaussian pdf, then the a posteriori pdf (Eq. (2)) can be described analytically by a Gaussian pdf, $\mathcal{N}(\tilde{\mathbf{m}},\tilde{\mathbf{C}}_{\mathrm{M}})$ (Tarantola and Valette, 1982a):

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_{\mathrm{M}}\mathbf{G}^t(\mathbf{G}\mathbf{C}_{\mathrm{M}}\mathbf{G}' + \mathbf{C}_{\mathrm{D}})^{-1}(\mathbf{d}_0 - \mathbf{G}\mathbf{m}_0) \tag{4}$$

$$\tilde{\mathbf{C}}_{\mathrm{M}} = \mathbf{C}_{\mathrm{M}} - \mathbf{C}_{\mathrm{M}}\mathbf{G}^t(\mathbf{G}\mathbf{C}_{\mathrm{M}}\mathbf{G}' + \mathbf{C}_{\mathrm{D}})^{-1}\mathbf{G}\mathbf{C}_{\mathrm{M}} \tag{5}$$

Note that Gaussian measurement errors and modelization errors combine through addition of the covariance operators, such that the combined covariance model is given by $\mathbf{C}_{\mathrm{D}} = \mathbf{C}_d + \mathbf{C}_T$. This allows accounting of Gaussian modelization errors directly as given in Eqs. (4) and (5) (Tarantola, 2005).

If $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{C}}_{\mathrm{M}}$ are available from Eqs. (4) and (5), then samples from the a posteriori pdf can be generated using e.g. Cholesky decomposition of the a posteriori covariance model, Eq. (5) in Le Ravalec et al. (2000).

Sampling the a posteriori pdf of a linear inverse Gaussian problem can also be performed using sequential Gaussian simulation without the need for explicitly computing $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{C}}_{\mathrm{M}}$ (Hansen et al., 2006). Hansen and Mosegaard (2008) extend this approach to work with direct sequential simulation. This allows a non-Gaussian a priori distribution of model parameters.

An alternative approach is to use kriging through error simulation (Journel and Huijbregts, 1978, p. 495), in a co-kriging formulation as proposed by Gloaguen et al. (2005a, 2005b) This approach may be faster than the methods based on sequential simulation, but is only valid for strictly Gaussian a priori models.

The above-mentioned methods rely on the fact that in a linear formulation, data can be seen as weighed averages of the model parameters. While not specifically making the link to inverse problems, such ideas has also been explored by Journel (1999) and Gómez-Hernández et al. (2005).

2.2. The non-linear Inverse problem

The linear and Gaussian assumptions considered above are convenient as they lead to computationally efficient algorithms. However, in reality the inverse problem is typically non-linear and the Gaussian assumption not valid. This may lead to severe artifacts in the inversion if the least-squares based approaches, as described above, are used. Instead one can use sampling techniques to sample the a posteriori pdf.

Rejection sampling: Perhaps the simplest method to sample the a posteriori pdf is the rejection sampler that can be implemented as follows:

1. Propose a model candidate from the a priori pdf, $\mathbf{m}_{pro}$.
2. Compute $L(\mathbf{m}_{pro})$.
3. Accept the proposed model as a realization of the a posteriori pdf with probability
$P_{acc} = L(\mathbf{m}_{pro})/L_{max}$ (6)

where L_{max} is the maximum value the likelihood function can obtain. Typically the value of L_{max} is not known and must be set to 1.

The only requirement for using the method is that one must be able to generate independent realizations of the a priori pdf and compute the corresponding likelihood. The collection of models accepted by the rejection sampling algorithm will be a sample of the a posteriori pdf. The main problem with the rejection sampler is that it is computationally very inefficient for anything but very low dimensional problems.

The extended Metropolis sampler: Mosegaard and Tarantola (1995) propose an extended version of the Metropolis algorithm (Metropolis et al., 1953; Hastings, 1970) that allows sampling the a posteriori pdf of an inverse problem with, in principle, arbitrary complex a priori information as given by Eq. (2). Using the classical Metropolis algorithm one must be able to evaluate the a posteriori probability $\sigma_M(\mathbf{m})$ and, hence, typically also the a priori probability, in order to evaluate Eq. (2).

The extended Metropolis algorithm differs from the classical Metropolis algorithm in that one neither need to evaluate the a posteriori probability $\sigma_M(\mathbf{m})$, nor the a priori probability $\rho_M(\mathbf{m})$ of a given model $\mathbf{m}$. If only an algorithm is present that can sample the a priori pdf and a method exist for evaluating the likelihood, $\rho_D(g(\mathbf{m}))$, then the extended Metropolis algorithm will sample the a posteriori pdf.

The extended Metropolis algorithm is a Markov Chain Monte Carlo method and can be implemented as a random walk in the space of a priori acceptable models as follows. If initially a realization of the a priori pdf is generated as $\mathbf{m}_{cur}$, and the associated likelihood $L(\mathbf{m}_{cur})$ is evaluated using Eq. (3), then the following algorithm will sample the a posteriori pdf:

1. In the vicinity of $\mathbf{m}_{cur}$, propose a new model candidate, $\mathbf{m}_{pro}$, consistent with the a priori model.
2. Compute $L(\mathbf{m}_{pro})$.
3. Accept the proposed model with probability $P_{acc} = \min([1, L(\mathbf{m}_{pro})/L(\mathbf{m}_{cur})])$.
4. If the proposed model is accepted, then the transition from $\mathbf{m}_{cur}$ to $\mathbf{m}_{pro}$ is accepted, and the proposed model becomes the current model, $\mathbf{m}_{cur} = \mathbf{m}_{pro}$. Otherwise, the random walker stays a location $\mathbf{m}_{cur}$ and $\mathbf{m}_{cur}$ counts again.

There are only two requirements for running the extended Metropolis algorithm: (1) One must be able to evaluate the likelihood function, Eq. (3). This is most often trivial, even if it may be computationally demanding, as it requires one to solve the forward problem and evaluate the corresponding data fit given the noise model. (2) One must be able to sample the a priori pdf such that aperiodicity and irreducibility is ensured (Mosegaard and Sambridge, 2002). In addition, it is preferable to be able to control the exploratory nature (often referred to as the step length) of the sampling algorithm, i.e. step 1 in the above algorithm, which is closely linked to the computational efficiency. See Mosegaard and Tarantola (1995) for details on the extended Metropolis algorithm.

The sequential Gibbs sampling algorithm provides such a general way to sample complex a priori models, with arbitrary step length ensuring aperiodicity and irreducibility (Hansen et al., 2012). Sequential Gibbs sampling can be used with any pdf that can be sampled using sequential simulation, which is the case for most of the statistical models developed in the geostatistical community over the last decades. The resampling strategy inherent in the sequential Gibbs sampler was initially proposed by Hansen et al. (2008), and subsequently Irving and Singha (2010) and Mariethoz et al. (2010) proposed similar methods. Hansen et al. (2012) demonstrate how the method is similar to an application of the Gibbs sampler and show that the method leads to a way of sampling the a priori pdf where aperiodicity and irreducibility is ensured.

3. SIPPI

SIPPI is a Matlab toolbox (SIPPI), compatible with Gnu Octave, that can be used to solve inverse problems in the formulation given by Eqs. (2) and (3) by allowing *S*ampling the solution to *I*nverse *P*roblems with complex A *P*riori *I*nformation.

In order to solve a probabilistic framed inverse problem as presented previously, one needs (at least) three ingredients: (1) a choice of an a priori model, (2) a choice of how to solve the forward problem, and (3) a choice of a noise model that describes the uncertainty of the observed data and the modelization error. Once these choices have been made one can solve the inverse problem using any of the applicable inversion methods.

SIPPI provides a generic approach to define the a priori model and the noise model in the form of the two data structures `prior` and `data`.

3.1. The a priori model

All information about the a priori model are defined in the Matlab structure called `prior`, which can specify any number of a priori type of models. For example an a priori choice of a 2D Gaussian velocity field can be specified in `prior{1}` and a 1D parameter describing a bias correction can be specified in `prior{2}`. Once the `prior` has been defined, a realization of the corresponding a priori pdf can be generated by calling

```
m = sippi_prior(prior);
```

`m` is a Matlab structure of the same size as `prior`. If three types of a priori models have been defined in `prior{1}`, `prior{2}`, and `prior{3}`, then the corresponding realizations will be stored in `m{1}`, `m{2}`, and `m{3}`. Considering the example above, `m{1}` will hold a realization of a 2D a priori model, while `m{2}` will hold a realization of a 1D a priori model. For the remainder of the text the index `im` will point to a specific number of a priori model, `prior{im}`.

A number of different types of a priori models can be selected using a `type` field to the `prior` data structure. The following four types of a priori models are available as part of SIPPI:

```
im=1;
prior{im}.type='GAUSSIAN';
prior{im}.type='FFTMA';
prior{im}.type='VISIM';
prior{im}.type='SNESIM';
```

Generalized Gaussian: `prior{im}.type=GAUSSIAN'` defines a 1D generalized Gaussian distribution

$$f_{gg}(m_0,\sigma,p) = \frac{p^{1-1/p}}{2\sigma\Gamma(1/p)} \exp\left(-\frac{1}{p}\frac{|m-m_0|^p}{\sigma^p}\right) \tag{7}$$

where p is the norm and σ is the variance. f_{gg} is symmetric around m_0, the a priori mean value. In the limit of $p \to \infty$ f_{gg} will define a uniform distribution. The following code defines a 1D Gaussian distribution with mean 10 and standard deviation 2:

```
im=1;
prior{im}.type='GAUSSIAN';
prior{im}.m0=10;
prior{im}.std=2;
```

If not set, the `norm` is by default set to 2. The following code defines a 1D close to uniform distribution in the interval [8,12]:

```
im=1;
prior{im}.type='GAUSSIAN';
prior{im}.m0=10;
```

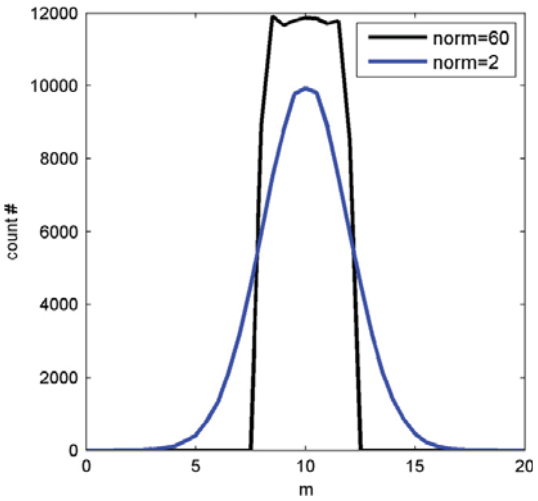


Fig. 1. Histogram of 100,000 unconditional realizations from a generalized Gaussian, GAUSSIAN type prior model with norm 60 and 2.

```
prior{im}.std=2;
prior{im}.norm=60;
```

A histogram of a sample of size 100,000 of these two 1D prior models is shown in Fig. 1.

The FFTMA, VISIM and SNESIM type priors all describe a 1D to 3D a priori model defined on a Cartesian grid, which is defined as (for a 3D case)

```
im=1;
prior{im}.prior.x=[0:1:10]; % X array
prior{im}.prior.y=[0:1:20]; % Y array
prior{im}.prior.z=[0:1:30] ; % Z array
```

For a 1D prior only `prior{im}.prior.x` needs to be defined, and for a 2D prior `prior{im}.prior.x` and `prior{im}.prior.y` need to be defined.

Both the FFTMA and VISIM type a priori models describe a multivariate Gaussian a priori pdf, which requires the specification of an a priori mean and covariance model. The a priori mean `m0` can be either a scalar, indicating a constant a priori mean model, or a matrix of the size of the a priori model, allowing for a varying a priori mean model. The model of spatial variability is defined by a, possibly anisotropic, covariance model (equivalent to a semivariogram model) given by the `Cm` (or equivalent the `Va`) field. The specification of the covariance model uses the same notation as used in Pebesma and Wesseling (1998). For example a multivariate Gaussian model defined by a 2D Spherical type covariance model with sill (or variance) 1, a maximum correlation length of 10 in the direction west to east (i.e. horizontal), and a perpendicular range (i.e. vertical) of 2.5 (hence an anisotropy factor of 0.25) and a mean of 10, is given by

```
prior{im}.m0=10;
prior{im}.Cm='1 Sph(10,90,0.25)';
```

FFT moving average: `prior{im}.type='FFTMA'` defines a spatially correlated multivariate Gaussian a priori model where a priori realizations are generated using the FFT Moving Average generator (FFTMA) (Le Ravalec et al., 2000). The FFTMA algorithm is very efficient for generating unconditional realizations from a multivariate Gaussian model. In addition it also allows separation of the random component field and the structural parameters that define spatial correlation. We will discuss the use of this feature in more details later.

A 2D FFTMA type a priori model, on a 200×100 grid, can for example be given by

```
im=1;
prior{im}.type='FFTMA';
prior{im}.prior.x=[0:.1:10]; % X array
prior{im}.prior.y=[0:.1:20]; % Y array
prior{im}.m0=10;
prior{im}.Va='1 Sph(10,90,.25)';
```

Fig. 2a shows a set of five realizations from this choice of a priori model.

VISIM : `prior{im}.type='VISIM'` defines a spatially correlated multivariate Gaussian a priori model where a priori realizations are generated using the VISIM algorithm (Hansen and Mosegaard, 2008). VISIM can run using sequential Gaussian simulation, in which case the model parameters are assumed to be normally distributed. It can also run using direct sequential simulation, which allows a (non-Gaussian) target distribution to be set that describes the a priori distribution of the model parameters, while at the same time ensuring that the a priori chosen mean and covariance will be honored.

An a priori model similar to the one described above for the FFTMA type prior, but with an a priori assumption of a bimodal distribution of model parameters can be given as

```
im=1;
prior{im}.type='VISIM';
prior{im}.prior.x=[0:1:10]; % X array
prior{im}.prior.y=[0:1:20]; % Y array
prior{im}.m0=10;
prior{im}.Va='1 Sph(10,90,.25)';
% target distribution
N=10,000;
prob_chan=0.5;
d1=randn(1,ceil(N*(1-prob_chan)))*.5+8.5;
d2=randn(1,ceil(N*(prob_chan)))*.5+11.5;
d_target=[d1(:);d2(:)];
prior{im}.target=d_target;
```

Fig. 3 shows a set of five realizations from this VISIM type of a priori model (a) without a specification of a target distribution and (b) using a target distribution. Once `[m,prior]=sippi_prior(prior)` has been called once, a data structure will be available as `prior{im}.V`, which allows access to all options available for running the VISIM algorithm. See Hansen and Mosegaard (2008) for more details on VISIM.

The FFTMA and VISIM type prior models only allow reproducing the first two moments of the distribution describing the spatial variability, the mean and the covariance (i.e. Gaussian variability between sets of two data points). Maximum entropy is implicitly assumed in higher order moments (Journel and Zhang, 2006). This is the reason why geological structures such as for example meandering channels cannot be reproduced by Gaussian statistics. To achieve this one can make use of statistical models based on higher order moments.

SNESIM : `prior{im}.type='SNESIM'` defines an a priori model based on a higher order statistical moments (a multiple point statistical model) describing spatial variability as inferred from a training image.

There are several methods that allow sampling from an a priori model defined by multiple point statistics. Here, we use the

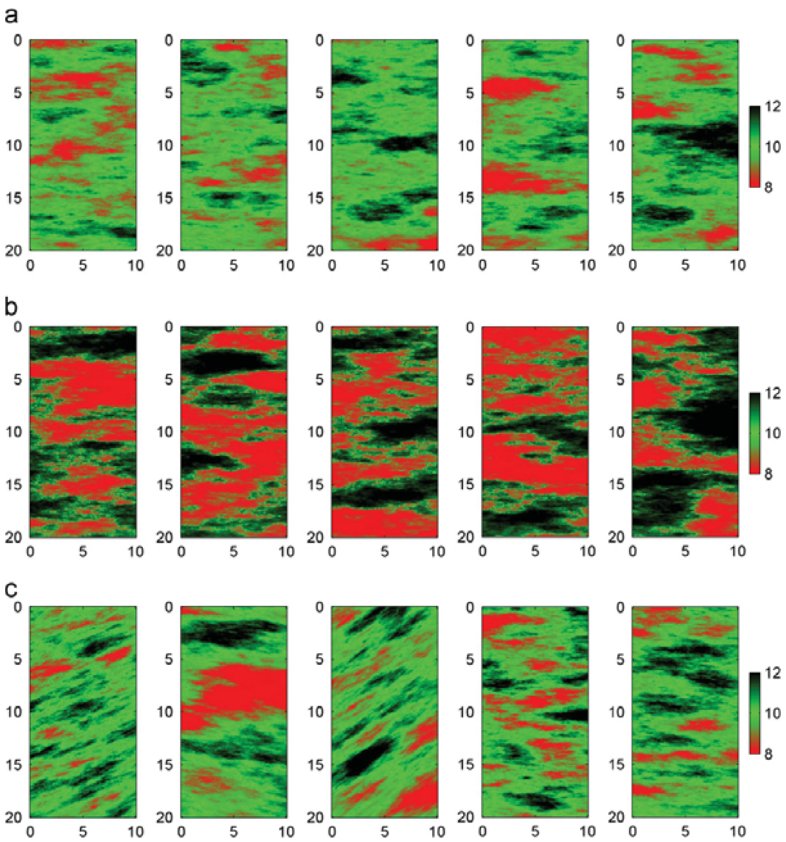


Fig. 2. Unconditional realizations from a `FFTMA` type priori model with (a) Gaussian distribution, (b) target distribution, and (c) random structural parameters (range and rotation).

SNESIM algorithm, originally developed by Strebelle (2000, 2002), and we make use of the implementation available in the SGeMS software package (Remy et al., 2008). It works by initially extracting a multiple point based statistical model from a training image. Then sequential simulation is used to generate realizations of this statistical model.

Optionally the `scaling` and `rotation` field can be specified. `prior{im}.scaling=2` scales the axis of the training image such that spatial structures appears twice as large. `prior{im}.rotation=45` rotates the training image 45° clockwise.

A 2D `SNESIM` type prior with the training image 'channels.ti' (Fig. 4) rotated 30° and scaled by a factor of 0.75, with two categories ('0' and '1'), and where the first category '0' reflect a model parameter value of 8, and the second category '1' reflect a value of 12, is given by

```
im=1;
prior{im}.type='SNESIM';
prior{im}.x=[0:.1:10];
prior{im}.y=[0:.1:20];
prior{im}.ti='channels.ti';
prior{im}.index_values=[0 1]; % optional
prior{im}.m_values=[8 12]; % optional
prior{im}.scaling=.75; % optional
prior{im}.rotation=30; % optional
```

Fig. 5 shows a set of five realizations from this choice of a priori model. Once `[m,prior]=sippi_prior(prior)` has been called, a data structure will be available as `prior{im}.S` which allow access to all options available for running the SNESIM algorithm as implemented in SGeMS. See Remy et al. (2008) for more details on setting up the SNESIM algorithm.

Distribution transform: A normal score transform can be defined for any of the Gaussian based a priori models, that allow the transformation of the normally distributed model parameters to any desired distribution, see e.g. Goovaerts (1997). It requires only that the user defines the 'target' distribution, in the form of a sample of the target distribution in the `d_target` field. For example a bimodal distribution with increased probability of values around 8.5 and 11.5, can be given by

```
N=10,000;
prob_chan=0.5;
d1=randn(1,ceil(N*(1-prob_chan)))*.5+8.5;
d2=randn(1,ceil(N*(prob_chan)))*.5+11.5;
d_target=[d1(:);d2(:)];
prior{im}.d_target=d_target;
```

Note that the number *N* here reflects the size of the sample generated and used to describe the target distribution in the `d_target` field, and can be chosen arbitrarily large. The larger the sample, the better the accuracy of reflecting a specific

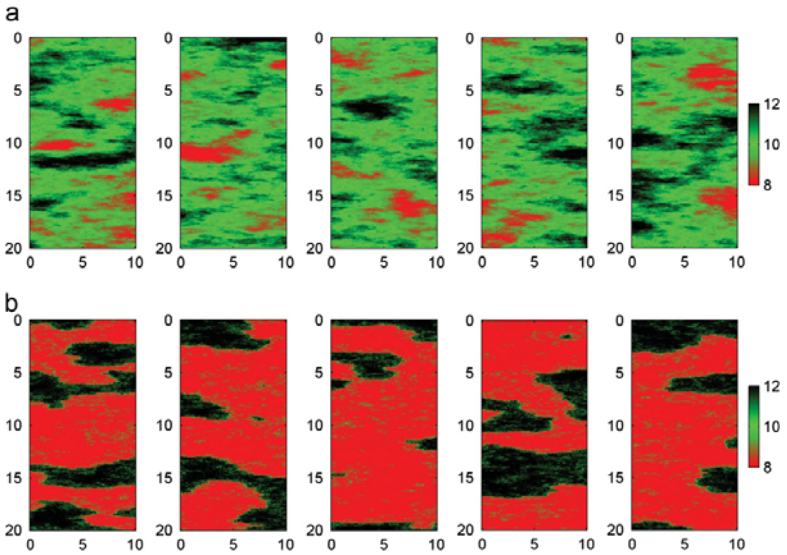


Fig. 3. Unconditional realizations from a VISIM type a priori model with (a) Gaussian distribution and (b) target distribution.

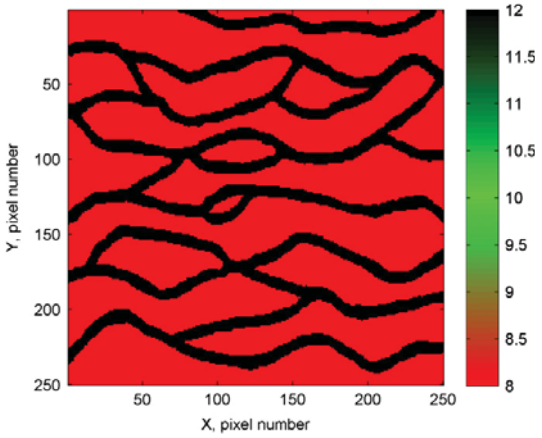


Fig. 4. Example of a training image for use with the SNESIM type a priori model.

distribution. An example of combining this distribution transform with the FFTMA type prior used to generate Fig. 2a is shown in Fig. 2b.

Note that when using the VISIM type prior one can use a target distribution directly, while ensuring that the chosen a priori covariance model is still honored. Using the distribution transform with the FFTMA prior will not preserve the properties of the a priori chosen covariance model.

Randomizing the model of spatial variability: As mentioned for the 'FFTMA' prior type model, the structural parameters that describe the a priori model covariance can be separated from the random number series that defines the random component. Therefore, all properties of the covariance model can be treated as model parameters, such as scaling and rotation. The properties of the model covariance can be perturbed independently of the random number series defining the random component (Le Ravalec et al., 2000).

In order to randomize a specific component of the covariance model, a GAUSSIAN type prior model needs to be defined for this component. The name of the specific prior model must be either range_1, range_2, or range_3 to define the range, or one of ang_1, ang_2, or ang_3 to define the rotation, and m0 to define the a priori mean, and sill to define the sill. In addition, one must set the prior_master field to point the prior model that define the prior for the corresponding FFTMA a priori model.

As an example, consider the FFTMA example used to generate Fig. 2a. To randomize the maximum correlation length to be close to uniform between 6 and 14, and randomize the primary rotation angle to be close to uniform between 40 and 130 degrees (from north) use

```
im=1;
prior{im}.type='gaussian';
prior{im}.name='range_1';
prior{im}.m0=10;
prior{im}.std=4;
prior{im}.norm=80;
prior{im}.prior_master=3;
im=2;
prior{im}.type='gaussian';
prior{im}.name='ang_1';
prior{im}.m0=90;
prior{im}.std=50;
prior{im}.norm=80;
prior{im}.prior_master=3;

im=3;
prior{im}.type='FFTMA';
prior{im}.prior.x=[0:1:10]; % X array
prior{im}.prior.y=[0:1:20]; % Y array
prior{im}.m0=10;
prior{im}.Va='1 Sph(10,90,.25)';
```

Fig. 2c shows an example of five realizations from such an a priori model.

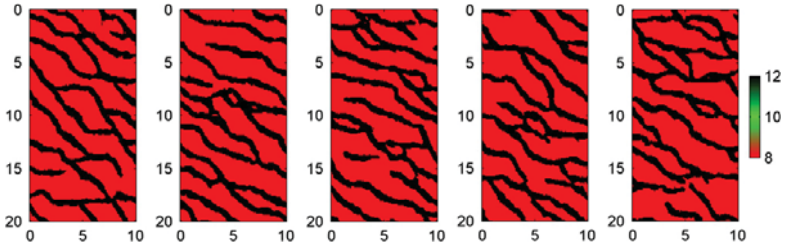


Fig. 5. Unconditional realizations from a SNESIM type a priori model.

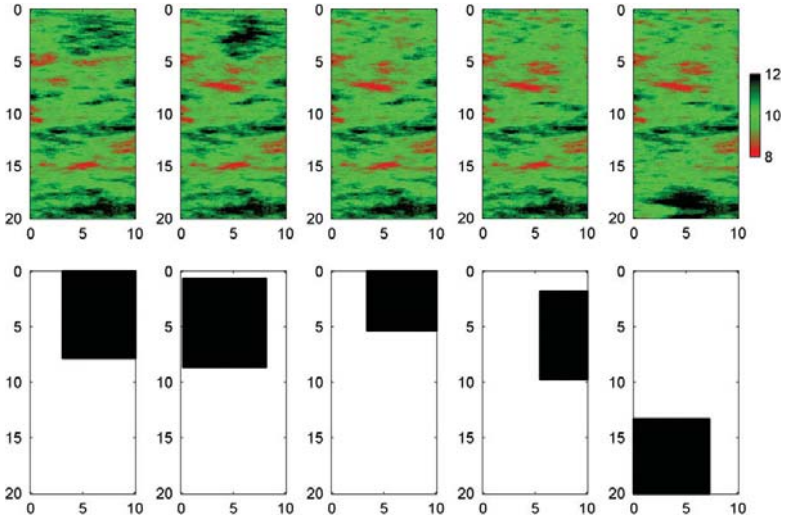


Fig. 6. (Top) Random walk using sequential Gibbs sampling with box type re-simulation, and the VISIM type a priori model. (Bottom) Black pixels indicate the model parameters that are simulated conditional to the value of the model parameters indicated by pixels.

3.1.1. A random walk in the a priori model space

To perform a random walk in the prior probability space, as needed by the extended Metropolis sampler, we make use of sequential Gibbs sampling (Hansen et al., 2012). An application of the sequential Gibbs sampler essentially amounts to selecting a subset, which can be any subset of model parameters, and simulate these conditional to the rest of the model parameters. The number of chosen model parameters in the subset controls the exploratory nature (i.e. step-length) of the sequential Gibbs sampler (which controls the degree of correlation between successive realizations), and hence the efficiency of the extended Metropolis sampler. All properties of the sequential Gibbs sampler are controlled by the `seq_gibbs` structure, which is a field in the `prior` data structure. Two different methods for selecting the subset of model parameters for conditional re-simulation have been implemented.

Box type subset: If `prior{im}.seq_gibbs.type=1`, then a line/rectangle/cube of model parameters (for the 1D, 2D and 3D case, respectively) is selected as the subset used for conditional re-simulation. The width of the box is defined by `prior{im}.seq_gibbs.step`. For example a box with dimension $2 \times 3 \times 4$ (in the units of the prior model considered - typically meters) is given by `prior{im}.seq_gibbs.step=[2 3 4]`. The center of the 'box' is chosen randomly

Randomly selected subset: If `prior{im}.seq_gibbs.type=2`, then a randomly selected number of the total number of model parameters is selected as the subset used for conditional resimulation. The number of data used for conditional re-simulation is given by `prior{im}.seq_gibbs.step`. If `prior{im}.seq_gibbs.step` is smaller than 1, it is interpreted as a percentage of the total number of model parameters.

As an example, five iterations of sequential Gibbs sampling can in SIPPI be performed using iterative calls to `sippi_prior` as

```
[m_current,prior]=sippi_prior(prior);
for i=1:5
[m_proposed,prior]=sippi_prior(prior,
m_current);
end
```

Figs. 6 and 7 show examples of using sequential Gibbs sampling with a box type selection and random type selection of model parameters for conditional re-simulation, respectively. The a priori model is in both cases the same as the one used to generate the unconditional realizations of Fig. 3. The options for the box type re-simulation are

```
prior{im}.seq_gibbs.type=1;
prior{im}.seq_gibbs.step=[4 4];
```

while the options for the random type re-simulation, with only 0.5% of the total number of model parameter used as conditional data for re-simulation, are

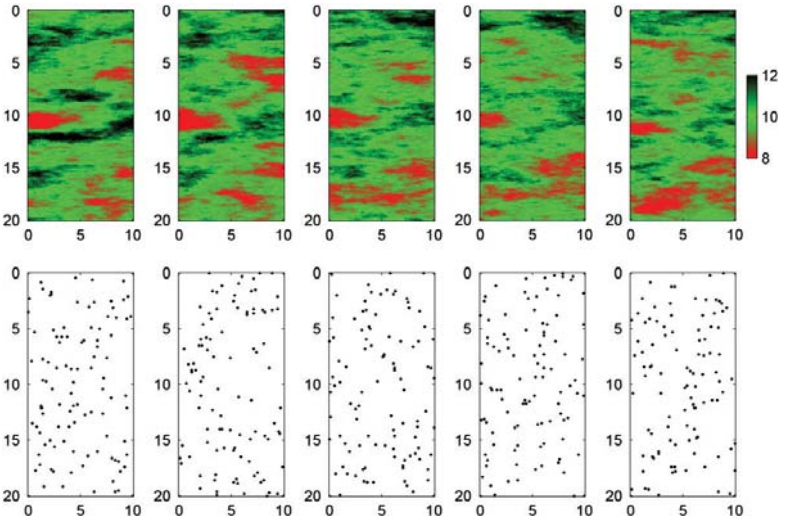


Fig. 7. (Top) Random walk using sequential Gibbs simulation with random choice of model parameters for resimulation, and the VISIM type a priori model. (Bottom) Black pixels indicate the model parameters that are simulated conditional to the value of the model parameters indicated by white pixels.

```
prior{im}.seq_gibbs.type=2;
prior{im}.seq_gibbs.step=0.995;
```

The sequential Gibbs sampler can be used with the FFTMA, VISIM, and SNESIM types a priori models. For the 1D GAUSSIAN type a priori model we use an alternate method. Given a current realization of the a priori model, a step length between 0 and 1 will generate a new realization of the prior, in the vicinity if the current realization. A step length of '0' indicates no change, while a step length of '1' will generate a new unconditional realization of the a priori model.

Fig. 8 shows the first 300 iterations when sampling the same a priori model as sampled in Fig. 1 using a step length of 0.25, `prior{im}.seq_gibbs.step=0.25`. After 100,000 iterations the histogram of the sampled model parameters resemble that of Fig. 1, and is therefore not shown here.

3.2. Data, data uncertainties, modelization errors and the likelihood function

Observed data must be given in the `data` data structure along with a description of the noise model. As for the `prior` structure, the `data` structure may consist of many types of data, where each data type number `id` is defined in the `data{id}` structure. Observed data are stored in the `d_obs` field. Uncorrelated uncertainty can be given either in the form of standard deviation, `d_std`, or variance, `d_var`. A simple data structure with such uncorrelated uncertainties can be given by

```
id=1;
data{id}.d_obs=[0 3 4]';
data{id}.d_std=[2 2 2]';
```

If the data uncertainties are uncorrelated, the noise model can be described by a generalized Gaussian model as defined in Eq. (7), if the norm of the generalized Gaussian is set by `data{id}.norm`. If not specified a Gaussian noise model (using a norm of 2) is chosen by default.

The noise model can also be given in the form of a correlated Gaussian model, for both the data noise, $\mathbf{C}_d$, and the modelization error, $\mathbf{C}_T$. The following will for example specify a correlated Gaussian noise model:

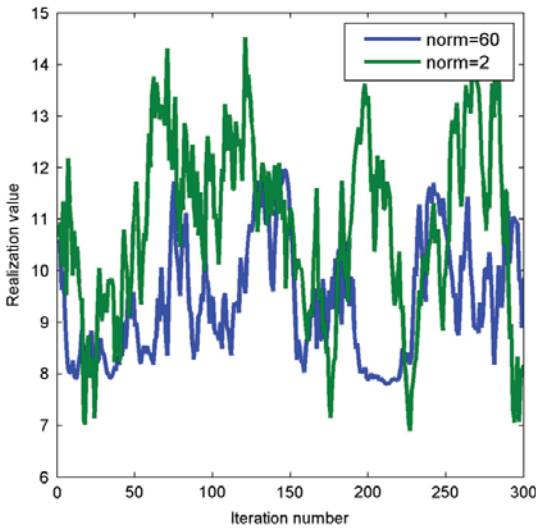


Fig. 8. The first 300 realizations from the GAUSSIAN type a priori model with a mean of 10, and a norm 60 and 2, respectively, using a step length of 0.25.

```
id=1;
data{id}.d_obs=[0 3 4]';
data{id}.Cd=[4 0 .1 ; 0 4 0 ; .1 0 4];
```

If a Gaussian model for the modelization error, $\mathcal{N}(\mathbf{d}_T,\mathbf{C}_T)$, is available it can be specified as

```
data{id}.dt=[0 -1 0]';
data{id}.Ct=[4 .1 .1 ; .1 4 .1 ; .1 .1 4];
```

where $\mathbf{d}_T$ is a bias correction.

One can choose to consider only a subset of the available data using the `i_use` field. To use for example only data number 1 and 3 use

```
id=1;
data{id}.d_obs=[0 3 4]';
data{id}.i_use=[1 3];
```

Once the data structure has been setup in `data`, the log-likelihood and the likelihood of a given data response *d* can be computed using

```
[log L, L, data]=sippi_likelihood(d,data);
```

3.3. The forward problem

The forward problem is naturally problem dependent, and to use SIPPI, the user needs to supply the solution to the forward problem, wrapped in the m-file `sippi_forward.m`.

The input to `sippi_forward.m` is the `forward`, `data` and `prior` Matlab structures. The `forward` structure can contain information on how to solve the forward problem. The output must be the data obtained by solving the forward problem, in the form of the data structure `d` which must be of the same length as the `data` structure, and each entry of `d{id}` must have the same size as `data{id}.d_obs`, or the size of `data{id}.i_use` if a data subset is specified.

As an alternative for providing `sippi_forward`, one can provide a generic name for the m-file solving the forward problem by setting `forward.forward_function`. Part 2 of this paper will provide an example of setting up `sippi_forward.m` (Hansen et al., this issue).

When the forward model has been setup, the process of generating an unconditional realization of the a priori model, `m`, followed by solving the forward problem and computing the likelihood of `m` can be done using

```
m=sippi_prior(prior);
d=sippi_forward(m,forward,prior,data);
log L=sippi_likelihood(d,data);
```

In the specific case where the forward relation is linear, the linear forward operator must be specified as the matrix *G*

```
forward.G
```

such that the forward problem can be solved using `d{1}=forward.G * m{1}`.

3.4. Sampling the a posteriori pdf

When the forward problem, `sippi_forward`, and the `prior`, `data`, and `forward` data structures have been defined, the a posteriori pdf can be sampled using the rejection sampler or the extended Metropolis sampler in the general non-linear case. In the linear Gaussian case, least-squares based inversion can be utilized.

3.5. Rejection sampling

Simple rejection sampling, using 30,000 iterations, of the a posteriori pdf can be performed using

```
options.mcmc.nite=30,000;
sippi_rejection(data,prior,forward,options);
```

By default the $L_{max}=1$, see Eq. (6). This can be manually changed by providing the `options.mcmc.Lmax`.

3.5.1. Metropolis sampling

All available a priori model types and noise models in SIPPI - work seamlessly as part of the extended Metropolis algorithm. The extended Metropolis sampling algorithm can be applied using

```
options=sippi_metropolis(data,prior,forward,options);
```

The `options` structure define some properties of how the Metropolis algorithm will run.

`options.mcmc.nite` determines the number of iterations of the extended Metropolis algorithm. `options.mcmc.i_sample` sets how often the current model is saved to disc, measured in the number of iterations. `options.mcmc.i_plot` sets the number of iterations between updating figures showing the progress of the algorithm. If any of these parameters are not set, then the following values will be chosen by default:

```
options.mcmc.nite= 30,000;
options.mcmc.i_sample= 500;
options.mcmc.i_plot: 50
```

Perturbation strategy: The choice of the number of model parameters to be perturbed in each iteration of the extended Metropolis algorithm can have large impact on its computational performance. By default a random type of model parameter is perturbed in each iteration. Thus if three types of a priori models have been specified in `prior{1}`, `prior{2}`, and `prior{3}`, the probability of perturbing each individual type of prior model in each iteration is 1/3. This default behavior can be changed by choosing a perturbation strategy. `options.mcmc.pert_strategy.i_pert` selects the number of a prior model types to perturb, and `options.mcmc.pert_strategy.i_pert_freq` set the relative frequency of each selected type of prior model. Thus, to perturb prior model 1 and 3 (but never model 2), such that prior model 3 is perturbed nine times as often as prior type 1, one could use

```
options.mcmc.pert_strategy.i_pert=[1 3];
options.mcmc.pert_strategy.i_pert_freq=[1 9];
```

Automatic adjustment of the exploration rate (step length): The exploratory nature of the Metropolis sampling algorithm, controlled by the 'step length', has large impact on its computational demands. A small step-length provides a dense local sampling, but the algorithm will use many iterations to move away from the initial point, i.e. a less exploratory algorithm. A large step length will lead to a very exploratory sampling algorithm that will not get trapped in local minima, but many models that are proposed will be rejected. Gelman et al. (1996) argue that a step-length leading to an acceptance rate in the Metropolis sampler of about 20–40% will lead to a good compromise between exploration and rejection rate. SIPPI allows automatic detection of the step length leading to an acceptance rate specified by `prior{im}.seq_gibbs.P_target`, using the method given by Cordua et al. (2012). Note that the Metropolis sampler will not sample the a posteriori pdf correct until the step-length is fixed, and unchanged. Therefore, one can set the number of initial iterations in which adjustment of the step length is allowed using `prior{im}.seq_gibbs.i_update_step_max`. After this, actual sampling of the a posteriori pdf will start, if the algorithm has reached burn-in. `prior{im}.seq_gibbs.i_update_step` sets the number of iterations between updating the step length. `prior{im}.seq_gibbs.step_min` and `prior{im}.seq_gibbs.step_max` determine the minimum and maximum allowed step length.

The default choice of the step length is to use infinitely long step-length, resulting in a prior sampler generating statistically independent realization of the prior in each iteration.

As an example, a preferred acceptance ratio of 0.3, adjusted in the first 1000 iterations, allowing step lengths in the interval 1–100 (using type 1 data subset), can be specified using:

```
prior{im}.seq_gibbs.type=1;
prior{im}.seq_gibbs.step_min=1;
prior{im}.seq_gibbs.step_max=100;
prior{im}.seq_gibbs.step=100;
prior{im}.seq_gibbs.i_update_step_max=1000;
prior{im}.seq_gibbs.P_target=0.3;
```

3.5.2. Linear Gaussian inverse problems

In the specific case where the forward problem is linear, and the a priori model Gaussian, as defined by the `VISIM` of `FFTMA` type a priori model, the a posteriori pdf can be sampled directly without the need for the Metropolis algorithm using

```
[m_reals,m_est,Cm_est]
  =sippi_least_squares(data,prior,forward,
n_reals,lsq_type);
```

`n_reals` sets how many a posteriori realizations, as output in `m_reals`, that are generated. `lsq_type` determines the method used to solve sample the a posteriori pdf. `m_est` and `Cm_est` are the a posteriori mean and covariance as given by Eq. (5), and are only available if least squares types of inversion is performed.

Three methods described previously are available to generate samples of the a posteriori pdf, and can be selected by setting the `lsq_type` argument when calling `sippi_least_squares`.

`lsq_type='lsq'` uses classical least-squares inversion where the complete Gaussian a posteriori pdf can be analytically derived in the form of a posteriori mean and covariance of Eqs. (4) and (5). Then Cholesky decomposition of the a posterior covariance is used to generated realizations of the a posteriori pdf.

`lsq_type='error_sim'` makes use of kriging simulation through error simulation to generate a sample of the a posteriori pdf (Journel and Huijbregts, 1978; Gloaguen et al., 2005a, 2005b; Hansen and Mosegaard, 2008).

`lsq_type='visim'` makes use of the VISIM algorithm for sampling the a posteriori pdf (Hansen and Mosegaard, 2008). The type of prior model must be chosen as a VISIM type prior model. If the target distribution is set as `prior{im}.target`, then VISIM runs as a direct sequential simulation algorithm. If it is not set, VISIM will run as a sequential Gaussian simulation algorithm.

4. Conclusions

A generic Matlab and Gnu Octave toolbox for sampling the a posteriori pdf of linear and non-linear inverse problems has been presented. Prior information about the model parameters can be described by any number of the following types of a priori models: (1) 1D arbitrarily distributed pdf, (2) 1D–3D multivariate Gaussian pdf as sampled using the FFTMA method, (3) 1D–3D multivariate Gaussian model as sampled using the VISIM algorithm (utilizing both sequential Gaussian simulation and direct sequential simulation), or (4) 1D–3D multiple-point based statistical models as sampled using the SNESIM algorithm.

For linear Gaussian inverse problems the a posteriori pdf can be sampled using (1) traditional least squares inversion combined with Cholesky decomposition of the a posteriori covariance, (2) sequential Gaussian simulation, (3) direct sequential simulation and (4) Gaussian simulation through error simulation.

For non-linear and non-Gaussian inverse problems the a posteriori pdf can be sampled using the rejection sampler or the extended Metropolis sampler. The computational efficiency of the extended Metropolis sampler can be controlled by using a flexible perturbation mechanism, based on sequential Gibbs sampling, allowing arbitrary long or short step length. The choice of the step length can optionally be automatized.

The combination of the FFTMA method with the extended Metropolis algorithm allows treating the properties describing the Gaussian a priori model, to be treated as model parameters, and thus inferred as part of the inversion.

Acknowledgment

We thank DONG for financial support. Interfacing geostatistical algorithms have been done using the mGstat toolbox, http://mgstat.sourceforge.net/. SIPPI source code can be downloaded from http://sippi.sourceforge.net/.

References

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

Gelman, A., Roberts, G., Gilks, W., 1996. Efficient metropolis jumping rules. In: Bernardo, J., Berger, K., Dawid, A., Smith, A. (Eds.), Bayesian Statistics, vol. 5. Clarendon Press, Oxford, pp. 599–608.

Gloaguen, E., Marcotte, D., Chouteau, M., 2005a. A non-linear tomographic inversion algorithm based on iterated cokriging and conditional simulations. In: Leuangthong, O., Deutsch, C. (Eds.), Geostatistics Banff 2004, vol. 1. Springer, pp. 409–418.

Gloaguen, E., Marcotte, D., Chouteau, M., Perroud, H., 2005b. Borehole radar velocity inversion using cokriging and cosimulation. Journal of Applied Geophysics 57 (4), 242–259.

Gómez-Hernández, J., Froidevaux, R., Biver, P., 2005. Exact conditioning to linear constraints in kriging and simulation. In: Leuangthong, O., Deutsch, C. (Eds.), Geostatistics Banff 2004, vol. 2. Springer, pp. 999–1005.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Applied Geostatistics Series. Oxford University Press.

Hansen, T., Cordua, K., Looms, M., Mosegaard, K. SIPPI: a Matlab toolbox for Sampling the solution to Inverse Problems with complex Prior Information: Part 2—Application to cross hole GPR tomography. Computers & Geosciences, http://dx.doi.org/10.1016/j.cageo.2012.10.001, this issue.

Hansen, T.M., Cordua, K.C., Mosegaard, K., 2012. Inverse problems with non-trivial priors—efficient solution through sequential Gibbs sampling. Computational Geosciences 16 (3), 593–611.

Hansen, T.M., Journel, A.G., Tarantola, A., Mosegaard, K., 2006. Linear inverse Gaussian theory and geostatistics. Geophysics 71 (6), R101–R111.

Hansen, T.M., Mosegaard, K., 2008. VISIM: sequential simulation for linear inverse problems. Computers and Geosciences 34 (1), 53–76.

Hansen, T.M., Mosegaard, K., Cordua, K.C., 2008. Using geostatistics to describe complex a priori information for inverse problems. In: Ortiz, J.M., Emery, X. (Eds.), VIII International Geostatistics Congress, vol. 1. Mining Engineering Department, University of Chile, pp. 329–338.

Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97.

Irving, J., Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resource Research 46.

Journel, A., Zhang, T., 2006. The necessity of a multiple-point prior model. Mathematical Geology 38 (5), 591–610.

Journel, A.G., 1999. Conditioning geostatistical operations to nonlinear volume averages. Mathematical Geology 31, 931–953.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press.

Le Ravalec, M., Noetinger, B., Hu, L.Y., 2000. The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulations. Mathematical Geology 32 (6), 701–723.

Mariethoz, G., Renard, P., Caers, J., 2010. Bayesian inverse problem and optimization with iterative spatial resampling. Water Resources Research 46 (11), W11530.

Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.

Mosegaard, K., Sambridge, M., 2002. Monte Carlo analysis of inverse problems. Inverse Problems 18 (3), 29–54.

Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research 100 (B7), 12431–12447.
Mosegaard, K., Tarantola, A., 2002. Probabilistic approach to inverse problems. In: Lee, W., Kanamori, H., Jennings, P., Kisslinger, C. (Eds.), International Handbook of Earthquake and Engineering Seismology, vol. 81A, pp. 237–265 (Chapter 16).
Pebesma, E.J., Wesseling, C.G., 1998. Gstat: a program for geostatistical modelling, prediction and simulation. Computers & Geosciences 24 (1), 17–31.
Remy, N., Boucher, A., Wu, J., 2008. Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press.
Strebelle, S., 2000. Sequential Simulation Drawing Structures from Training Images. Ph.D. Thesis, Stanford University.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34 (1), 1–20.
Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM.
Tarantola, A., Valette, B., 1982a. Generalized nonlinear inverse problems solved using the least squares criterion. Reviews of Geophysics and Space Physics 20 (2), 219–232.
Tarantola, A., Valette, B., 1982b. Inverse problems = quest for information. Journal of Geophysics 50 (3), 150–170.

Appendix A15:

SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information Part 2—Application to crosshole GPR tomography

Authors:

Thomas Mejer Hansen, Knud Skou Cordua, Majken Caroline Looms, and Klaus Mosegaard

Published in:

Computers & Geosciences

Computers & Geosciences 52 (2013) 481–492

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information
Part 2—Application to crosshole GPR tomography

Thomas Mejer Hansen [a,*], Knud Skou Cordua [a], Majken Caroline Looms [b], Klaus Mosegaard [a]

[a] Technical University of Denmark, Center for Energy Resources Engineering, DTU Informatics, Asmussens Alle, Building 305, DK-2800 Lyngby, Denmark
[b] University of Copenhagen, Department of Geography og Geology, Øster Voldgade 10, DK-1350 København K, Denmark

ARTICLE INFO

Article history:
Received 22 June 2012
Received in revised form
7 September 2012
Accepted 10 September 2012
Available online 12 October 2012

Keywords:
Inversion
Non-linear
Tomography
Sampling
A priori
A posteriori

ABSTRACT

We present an application of the SIPPI Matlab toolbox, to obtain a sample from the a posteriori probability density function for the classical tomographic inversion problem. We consider a number of different forward models, linear and non-linear, such as ray based forward models that rely on the high frequency approximation of the wave-equation and 'fat' ray based forward models relying on finite frequency theory. In order to sample the a posteriori probability density function we make use of both least squares based inversion, for linear Gaussian inverse problems, and the extended Metropolis sampler, for non-linear non-Gaussian inverse problems. To illustrate the applicability of the SIPPI toolbox to a tomographic field data set we use a cross-borehole traveltime data set from Arrenæs, Denmark. Both the computer code and the data are released in the public domain using open source and open data licenses. The code has been developed to facilitate inversion of 2D and 3D travel time tomographic data using a wide range of possible a priori models and choices of forward models.

1. Introduction

Tomographic inversion is used in many research fields such as geophysics and medical imaging. With this technique, images of an unknown 3D object can be obtained based on indirect observations from outside of the object. One such example is travel time inversion that can for example be used to map the internal velocity structure of the earth, based on recordings of the arrival times of certain seismic phases generated as part of e.g. an earthquake. Another example of a tomographic data set, is that obtained by measuring the travel time delay of a seismic or electromagnetic wave traveling between a source and a receiver. Given such a set of observed travel time data the tomographic inverse problem consists of inferring information about the velocity around and in-between the sources and receivers. It is this latter problem that we will address here using the SIPPI toolbox, which is a Matlab toolbox for sampling the solution to inverse problems with complex a priori information (Hansen et al., this issue).

We will specifically address the problem of first arrival travel time inversion using crosshole ground-penetrating radar (GPR) data. Such travel time data are sensitive to the subsurface variations in electromagnetic wave velocity, that is related to the dielectric permittivity, which is strongly influences by water moisture (Topp et al., 1980). Inversion of such travel time data thus has the potential to map subsurface moisture content.

For linear or weakly non-linear inverse problems least squares based methods are widely applied. Deterministic least squares methods is presented by e.g. Menke (1989), while a probabilistic approach is given by e.g. Tarantola and Valette (1982) and Tarantola (2005).

A probabilistic approach to linear travel time tomography, based on sequential simulation, was proposed by Hansen et al. (2006) and Hansen and Mosegaard (2008) who utilized the equivalence of classical least squares inversion (e.g. Tarantola and Valette, 1982) and kriging (e.g. Journel and Huijbregts, 1978). An application of this approach to crosshole georadar data is given in Nielsen et al. (2010). A related method based on kriging through error simulation (Journel and Huijbregts, 1978), equivalent with the probabilistic least squares approach, was proposed and applied to cross hole GPR tomographys by Gloaguen et al. (2005a,b). Recently this approach was applied for inversion of an anisotropic velocity field (Giroux and Gloaguen, 2012). These methods are only strictly valid for linear inverse problems, and rely on an inherent assumption of Gaussian statistics describing both the noise model and the a priori model. Specifically the a

* Corresponding author. Tel.: +45 45253086; fax: +45 45882673.
E-mail addresses: tmeha@imm.dtu.dk (T.M. Hansen), kcor@imm.dtu.dk (K.S. Cordua), mcl@geol.ku.dk (M.C. Looms), kmos@imm.dtu.dk (K. Mosegaard).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.10.001

priori model must be given in form of a Gaussian a priori model defined by a mean and a covariance model. Choosing such a Gaussian prior model may not be trivial. A number of methods have been developed to estimate this model prior to inverting the data (Asli et al., 2000; Hansen et al., 2008a; Irving et al., 2009; Looms et al., 2010).

For examples of least squares based deterministic tomographic inversion of GPR cross hole data see e.g. Irving et al. (2007) and Dafflon et al. (2011). Examples of stochastic inversion is presented for inversion of time lapse cross hole 1D travel time data by Scholer et al. (2012) and 2D time lapse electrical resistivity data by Irving and Singha (2010). Hansen et al. (2008b) demonstrate an application of the extended Metropolis sampler (Mosegaard and Tarantola, 1995) to a non-linear cross hole tomographic problem, where the a priori model is non-Gaussian and defined by any geostatistical method.

Here we will demonstrate the use of the SIPPI Matlab toolbox for solving the crosshole traveltime tomography inverse problem in a probabilistic framework. Initially we will briefly describe the theory describing different linear and non-linear solutions to the forward problem of computing the travel time delay between a propagating wave traveling between a source and a receiver. Then we will demonstrate how these forward models can be utilized with SIPPI. We will then make use of a reference data set obtained at Arrenæs, North Sealand, Denmark, to demonstrate all the inversion methods available in SIPPI, such as classical least squares estimation and simulation, and sampling methods such as the rejection sampler and the extended Metropolis sampler, see Hansen et al. (this issue).

2. Theory, first arrival travel time computation

The travel time delay of a propagating wave between a source and a receiver can be defined in a number of ways. We will consider methods based on the eikonal equation, 1st order sensitivity kernels and the Born approximation.

2.1. The eikonal equation

The eikonal equation describes the arrival time along a closed curve, $u(\mathbf{x})$, traveling with the speed defined by the velocity field, $m(\mathbf{x})$ (Sethian and Popovici, 1999)

$$|\nabla u(\mathbf{x})|\, m(\mathbf{x}) = 1 \tag{1}$$

Solving Eq. (1) allows locating the travel time, d, between a source and a receiver along the closed curve. To solve the eikonal equation we make use of an efficient implementation of the multistencil fast marching method proposed by Hassouna and Farag (2007), and made available by Dirk-Jan Kroon[1] under an open source license. This forward model is non-linear and, as the eikonal equation corresponds to a high frequency approximation to the wave equation. Therefore, it is often referred to as the high frequency ray approximation.

2.2. Forward models based on 1st order sensitivity kernels

The travel time d between a source and a receiver can be given by

$$d = \int G(\mathbf{x}) \frac{1}{m(\mathbf{x})}\, d\mathbf{x} \tag{2}$$

where $m(\mathbf{x})$ is the velocity field in which the signal travels. $G(\mathbf{x})$ is the sensitivity kernel that describes the sensitivity of each model parameter (within the Fresnell zone) to the travel time. $G(\mathbf{x})$ can be computed under a wide range of assumptions and thus defines the forward problem of computing the travel time delays in different ways.

2.2.1. Ray based forward model

Using the high frequency approximation to the wave equation results in a sensitivity kernel $G(\mathbf{x})$ that can be described by a ray connecting the source and receiver. Hence, this kernel can be obtained by solving the eikonal equation, which provides the fastest possible forward model. We will refer to this type of forward model as ray based.

2.2.2. Fat ray based forward model

Using a finite frequency (band limited) approximation to the wave equation leads to a sensitivity kernel where the sensitivity of the travel time delay also appears in a zone around the fastest ray path. A number of works have defined sensitivity kernels based on geometrical rules assigning sensitivity within the first Fresnel zone. Forward models based on these types of kernels will be referred to as fat ray based forwards (Husen and Kissling, 2001; Jensen et al., 2000).

2.2.3. Born based forward model

The Born approximation to the wave equation (considering only 1st order scattering) is an exact analytical expression for the sensitivity kernel for a point source, which can be derived for both seismic (Dahlen et al., 2000; Spetzler and Snieder, 2004; Marquering et al., 1999; Liu et al., 2009) and electromagnetic wave propagation (Buursink et al., 2008). The Born approximation also leads to a sensitivity kernel with sensitivity outside the ray approximation (i.e. a fat ray). The Born approximation is only strictly valid for a homogeneous velocity field, but have in practice been used also when the velocity field has relatively small velocity contrasts. For large velocity contrast this method becomes unstable and cannot be used. Forward models based on the Born approximation will be referred to as Born based forward models.

3. Cross hole GPR tomography at Arrenæs

As a case study we will demonstrate the capabilities of SIPPI for solving tomographic inverse problems. The implementation is generally applicable for travel time based tomographic problems, but here we will apply the toolbox to a cross hole GPR tomographic problem.

Initially we will present a 3D data set. Then we will demonstrate how the the different types of forward models have been implemented in `sippi_forward_traveltime` for easy utilization as part of SIPPI. Finally we demonstrate the use of SIPPI to solve the GPR cross hole tomography inverse problem using both linear and non-linear forward models, and simple and more complex a priori models.

3.1. Data: 3D GPR crosshole traveltime data from Arrenæs

As a reference data set we consider a 3D tomographic data set recorded as part of a ground penetrating radar (GPR) cross borehole survey at Arrenæs, North Sealand, Denmark. The data set we use here is identical to data presented by Looms et al. (2010), and is here made available in the public domain.

The observed data are first arrival times of electromagnetic waves propagating from a source location in one borehole to a receiver location in another borehole. Thus, the forward problem consists of estimating the travel time delay caused by the subsurface velocity field, given the recording geometry. The inverse

[1] http://www.mathworks.com/matlabcentral/fileexchange/24531-accurate-fast-marching.

problem is then to infer information about the subsurface velocity structure.

The subsurface at Arrenæs consists mostly of sand, with various degree of coarseness. The velocity of the subsurface is believed to represent natural moisture content. The lower the velocity the higher the moisture content (Topp et al., 1980).

Fig. 1 shows the relative position of four boreholes, AM1, AM2, AM3, and AM4. Tomographic travel time delay have been recorded between boreholes AM1–AM3 and AM2–AM4, respectively. The locations of the source and receiver positions down through the boreholes are shown in Fig. 1 and is marked by red dots in two of the boreholes. Note that the coloured ray like structure in Fig. 1 reflect the high frequency ray kernel related to a constant velocity model. The colors of each ray reflect the average velocity along each of the rays, and can be used as a rough indicator of the subsurface velocity structure.

Data are available as ASCII and binary Matlab formatted files for both the two 2D data sets, `AM13_data` and `AM24_data`, and the combined 3D data set, `AM1234_data` that combines the data sets `AM13_data` and `AM24_data`.

The Matlab mat files contain the location of the sources and receivers in the `S` and `R` variables. Observed data is in the `d_obs`

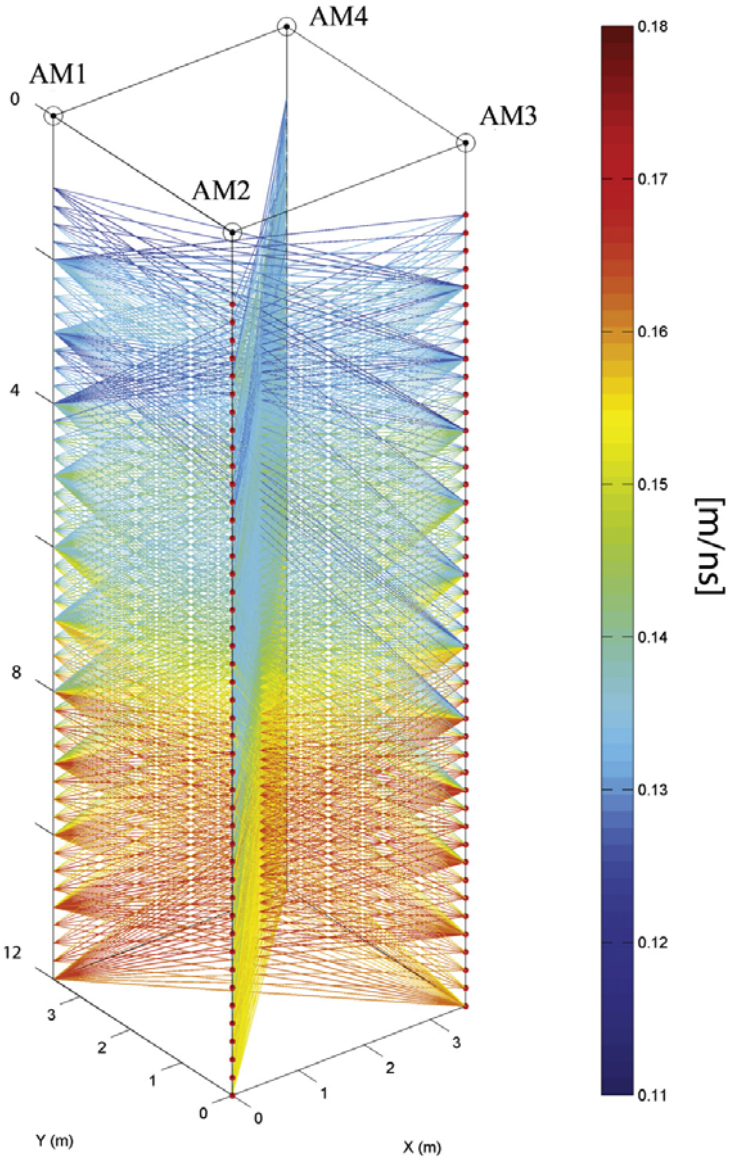


Fig. 1. Apparent ray coverage (using the linear high frequency approximation). The color of each ray reflects the apparent average velocity along the ray path. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

variable and the associated uncertainty (in form of the standard deviations) is in the d_std variable. A covariance model describing static like errors related to cross borehole GPR data, as given by Cordua et al. (2009), is available in the Ct variable.

3.2. The forward model—traveltime computation

As described in Hansen et al. (this issue), the only problem dependent part of using SIPPI is the implementation of a solution to the forward problem. We have implemented the m-file sippi_forward_traveltime that can be used to solve the forward problem of computing the travel time delay between a set of sources and receivers. All properties relating to solving the forward problem is defined in the forward Matlab structure. The output is the data structure d

```
[d]=sippi_forward_traveltime(m,forward,prior,data);
```

To make this solution of the forward problem available for the various inversion algorithms available in SIPPI, one can either implement an m-file sippi_forward that simply calls sippi_forward_traveltime, or one can specify the m-file to be used for solving the forward problem directly using forward.forward_function='sippi_forward_traveltime'. Note that this m-file and the specification of the forward structure is specific to the tomographic travel time inverse problem, while all other parts of the SIPPI toolbox are applicable to inverse problems in general.

Source and receiver geometry. The locations of the sources and receivers must be provided in the forward.sources and forward.receivers fields. Both the sources and receivers must point to a matrix with a number of rows equal to the number of rows (i.e. number of data) of data{id}.d_obs, and a number of columns reflecting the dimension of the prior model. For example, two sets of sources and receivers defined in 3D could be given by

```
forward.sources=[1 1 5; 1 1 10];
forward.receivers=[5 5 5; 5 5 10];
```

Forward model. Four types of forward models are available through sippi_forward_traveltime by specifying the forward.type field to one of eikonal, ray, fat, or born.

forward.type='eikonal' defines a forward model based on the solution to the eikonal equation, Eq. (1). This forward model is non-linear.

The other three available forward model types, ray, fat, and born, refer the the ray, fat and Born based sensitivity kernels presented earlier. When sippi_forward_traveltime is called using any of these types of forward models, a matrix operator, reflecting the choice of forward model, is computed as forward.G.

One can choose either a linear or non-linear formulation for solving such forward problems by specifying the forward.linear field. By default a non-linear formulation is assumed, such that forward.linear=0. This cause forward.G to be recalculated for each call to sippi_forward_traveltime. Different velocity models will result in different sensitivity kernels, and hence different forward operators, forward.G. Therefore, the forward problem is non-linear.

One can also choose a linear formulation, using forward.linear=1. In this case forward.G is only computed once, when sippi_forward_traveltime is called for the first time, and hence any subsequent calls to solve the forward model requires only a fast matrix multiplication. One can provide a velocity model for which the sensitivity kernel will be computed using forward.linear_m. If this is not specified the sensitivity kernel will be computed for the a priori mean model, given in prior{1}.m0.

forward.type='ray' selects the high frequency ray approximation presented earlier. This type of forward model is based on the same high frequency assumption as the eikonal type forward model. The difference is that here the forward operator forward.G is explicitly computed, which allows for a very fast forward model using forward.linear=0. If one would consider using the ray type forward model in a non-linear formulation, we suggest to use the eikonal type of forward model instead, which provides similar results but is computationally much more efficient. Used in the linear formulation this type forward model resemble the 'straight ray' approximation, as the the travel delay is due to the travel time delay along straight ray path that connects the source and receivers. The 'rays' on Fig. 1 reflect such a linear 'ray' type forward model.

forward.type='fat' selects a finite frequency (band limited) approximation to the wave equation, where the travel time delay i sensitive to a zone around the fastest ray path. Specifically the fat type forward model uses the empirical description of the travel time sensitivity kernel as proposed by Jensen et al. (2000), which is based on 1st order Fresnel zone sensitivity. The fat type forward model can be used both as linear and non-linear forward model.

forward.type='born' selects a forward model based on the Born approximation as presented earlier. Here we will make explicit use of the formulation of the sensitivity kernels given by Buursink et al. (2008). The born type forward model is only strictly valid for a homogeneous velocity field, but have in practice been used also when the velocity field has relatively small velocity contrasts. For large velocity contrasts this method becomes unstable and should not be used.

Using either forward.type='fat' or forward.type='born' the width of the sensitivity around the ray path, is related to the frequency of the propagating wave. Therefore, this frequency must be set as forward.freq. The frequency must be specified in the inverse unit of the observed travel time data given in data{id}.d_obs.

As an example of choosing the fat type forward model in a non-linear formulation using a wavelet frequency of 0.1 GHz, where traveltime data is measured in nanoseconds, is

```
forward.type='fat';
forward.freq=0.1;
forward.linear=0;
```

3.3. Solving the inverse problem

Having defined the forward problem, we will demonstrate the methods available in SIPPI for solving the inverse tomographic problem.

3.3.1. 2D non-linear inversion—AM13

Initially we will consider the 2D traveltime data set, AM13, recorded between well AM1 and AM3, using a simple Gaussian type a priori model. 702 travel time data and the position of associated source and receiver locations is available in the Matlab file AM13_data.mat. To use SIPPI, the forward, data, and prior structures need to be defined.

Setting up the forward *structure*. We use the high frequency ray approximation, in form of the eikonal type forward model, such that the forward data structure can be setup using

```
D=load('AM13_data.mat');
forward.sources=D.S;
forward.receivers=D.R;
forward.type='eikonal';
```

Setting up the `data` *structure*. The high frequency approximation, assumed using the eikonal solution, will always provide the fastest travel time between a source and a receiver, and always faster than the travel time of a wave with a finite finite frequency in a inhomogeneous velocity field. Therefore, we allow for a small modelization error, `Ct`, in form of a constant correlated Gaussian error of 1 ns^2 between all data. This will allow a small bias correction (the same for all data observations) to account for the relative fast travel times caused by the use of the high frequency forward model. The data in form of 702 observed traveltimes, `d_obs`, and associated uncorrelated uncertainties, `d_std` (of 0.7 ns), is available in the Matlab file `AM13_data.mat`. The `data` structure can be setup as

```
D=load('AM13_data.mat');
id=1;
data{id}.d_obs=D.d_obs;
data{id}.d_std=D.d_std;
data{id}.Ct=1;% modelizationerror
```

SIPPI allows using only a subset of the available data, which can be useful to test a certain setup relatively fast. The number of data considered is given by `data{id}.i_use`. To use every 20th data one could use `data{id}.i_use=20:20:702`. If not set it is automatically set to all data. In the current case this will be `data{id}.i_use=1:1:702`.

Setting up the `prior` *structure*. Looms et al. (2010) demonstrate a method for inferring the structural parameters of a Gaussian type a priori model. They tested their method on the data we use here and find an optimal a priori model for profile AM13 and AM24 independently. Initially we will make use of the same a priori model for both profile AM13 and AM24 and, therefore, based on the findings in Looms et al. (2010), we choose to use a Gaussian type a priori model as defined by a Spherical type covariance model with an isotropic covariance model with a range of 6m, a variance of 0.0003 m^2/ns^2, and a mean of 0.145 m/ns. We make use of the `FFTMA` type a priori model. The complete definition of the a priori model can then be given as

```
im=1;
prior{im}.type='FFTMA';
prior{im}.name='Velocity(m/ns)';
prior{im}.m0=0.145;
prior{im}.Va='.0003Sph(6)';
prior{im}.x=[-1:.2:6];
prior{im}.y=[0:.2:13];
```

A sample of the corresponding a priori model can then be generated and visualized using `sippi_plot_prior(prior)` as shown in Fig. 2a.

Sampling the a posteriori pdf using the extended Metropolis algorithm. Given the `forward`, `prior`, and `data` structures the extended Metropolis algorithm can be setup and run using e.g.

```
options.mcmc.nite=500000;
options.mcmc.i_plot=200;
options.mcmc.i_sample=250;
sippi_metropolis(data,prior,forward,options);
```

This will cause the extended Metropolis sampler to run for 500,000 iterations. The currently visited model will be saved to disk for every 250 iterations as specified by `options.mcmc.i_sample`

As the Metropolis algorithm is running, some properties are visualized for every `options.mcmc.i_plot` iterations, such as the currently accepted model, the step length for each prior type, and the log-likelihood curve. Such figures are often useful in the phase where the properties of the Metropolis algorithm are selected, prior to performing a full sampling.

Fig. 3 shows the log-likelihood value as function of the iteration number. The Metropolis algorithm has reached burn-in after about 2000 iterations as it reaches the plateau of log-likelihood values of approximately −90.

Recall that the way the sequential Gibbs sampler works, is controlled by the `prior{1}.seq_gibbs` structure, Hansen et al. (this issue). Here we make use of the default settings

```
prior{1}.seq_gibbs.i_update_step=50
prior{1}.seq_gibbs.i_update_step_max=1000
prior{1}.seq_gibbs.P_target=0.3000
```

This means that the step length of the Metropolis sampler is adjusted for every 50 iterations with the goal of achieving an acceptance rate of 0.3. After 1000 iterations the step length will be kept constant.

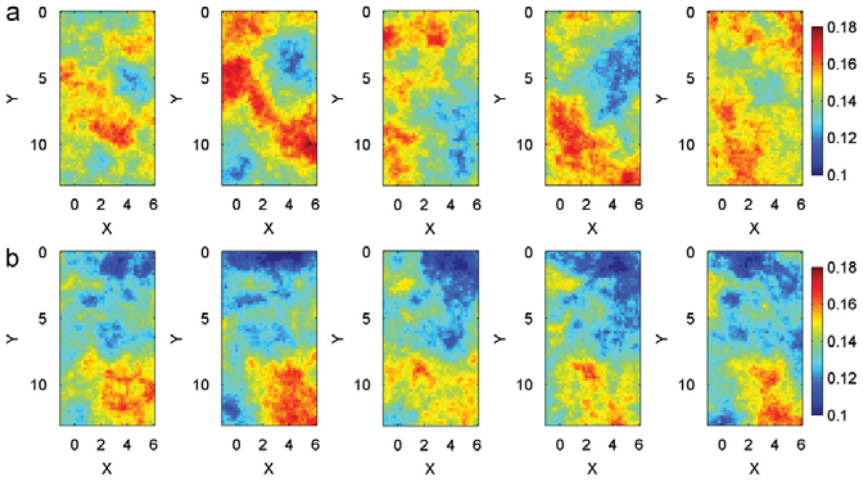


Fig. 2. Five realizations from the (a) a priori model and (b) a posteriori pdf considering data set AM13.

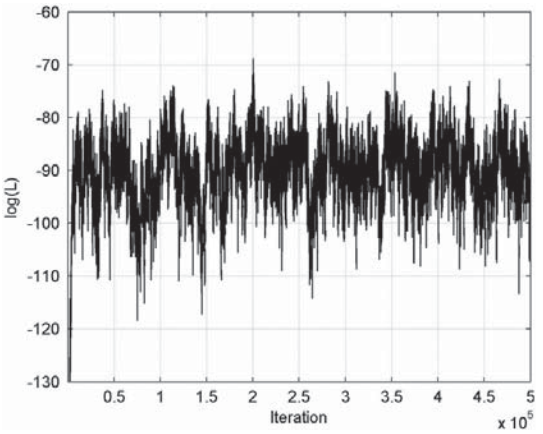


Fig. 3. Likelihood as a function of iteration number.

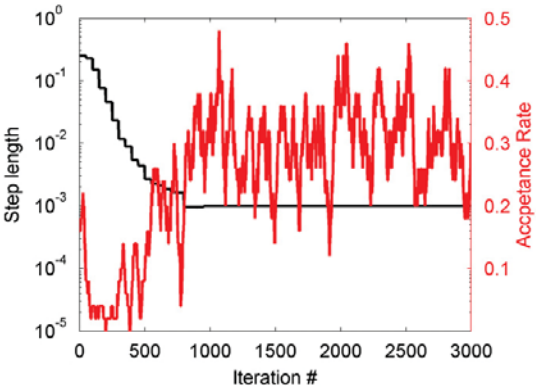


Fig. 4. Step length and acceptance rate of the Metropolis algorithm during the first 3000 iterations.

Fig. 4 shows the step length of the sequential Gibbs sampler as well as the acceptance rate in the first 3000 iterations. In the first 1000 iterations the step length is allowed to vary, and after 1000 iterations the step length stabilize around 10^{-3}. Initially the acceptance rate is about 0.2. Then it decreases rapidly until the step length is gradually adjusted, such that the acceptance rate ends up around 0.3, just as requested by `prior{1}.seq_gibbs.P_ target`. Recall that while the step length is being changed, and until the Metropolis algorithm has reached burn-in, the a posteriori pdf is not sampled (Cordua et al., 2012).

Fig. 2b shows five independent realizations of the a posteriori pdf, obtained after the Metropolis algorithm has reached burn-in. Comparing the realizations of the a posteriori pdf to the realization of the a priori pdf (Fig. 2a) reveals that the apparent scales and spatial structures visible in the a priori realizations are also present in the a posteriori realizations. The location of these structures is not resolved in the a prior realizations. But in the a posteriori realizations it is clear that relative high velocity structures dominate in the lower right corner while areas of lower velocity dominate the upper part of the model. Features such as these, that appear on many realizations of the a posteriori pdf are well resolved features (Mosegaard, 1998).

Once the extended Metropolis sampler has finished a number of plots for quality control can be generated using `sippi_plot_posterior`. First a figure visualize a sample of the a posteriori pdf, as in Fig. 2a. Second, a figure shows the acceptance ratio and step length as a function of iteration number, as in Fig. 4. Third, a figure shows the distribution of data residuals, i.e. the difference between observed and simulated travel time data, corresponding to number a realizations of the a posteriori pdf, as in Fig. 5. Note how the distribution is very close to Gaussian, as defined in the noise model. Note also how the distribution is not centered around 0 ns, but has a mean value (i.e. a bias) of about -1.5 ns. This is due to allowing a constant modelization error of 1 ns^2, that was applied in order to account for the use of the eikonal type forward model, that will always provide the fastest possible travel time between a source and a receiver. This is correctly reflected in the negative bias correction.

Finally `sippi_plot_posterior` provides a figure that illustrates the correlation coefficient of the currently accepted model in the last iteration to any of the other models sampled from the a posteriori pdf. This is used to estimate the number of iterations between independent realizations of the a posteriori pdf, e.g. Cordua et al. (2012). An example generated for the present example, is shown in Fig. 6. The correlation coefficient between the current model at iteration 500,000 and the models close to iteration number 500,000 is close to 1, and such models are not statistically independent. However, in a number of iterations away from the last considered model, the correlation coefficient decreases, until it reached a level of around 0.7. We use this level of the correlation coefficient to determine the approximate number of iterations between independent realizations of the a posteriori pdf obtained by the Metropolis algorithm. For the present case this was estimated to be around 10000 iterations between independent realizations.

Sampling the a posteriori pdf using the rejection sampler. Sampling the a posteriori pdf for the tomographic inverse problem using rejection sampling, can in principle be performed using

```
options.mcmc.nite=500000;
sippi_rejection(data,prior,forward,options);
```

The maximum a posteriori likelihood L_{max} is set to 1, if not, as here, specifically set using `options.mcmc.Lmax`, see Hansen et al. (this issue). Fig. 7 (green bars) shows a histogram of the likelihood of all the a posteriori accepted models using the extended

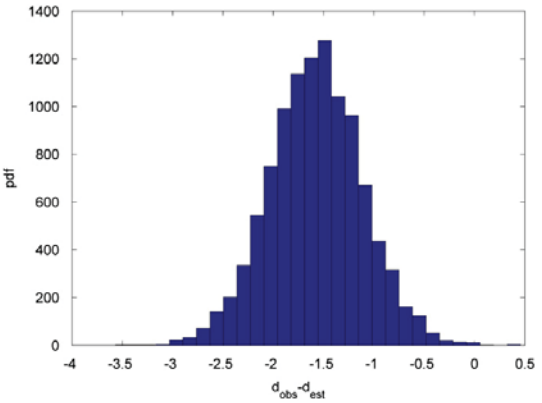


Fig. 5. Distribution of the difference between observed traveltime data and the traveltime data associated to 10 realizations of the a posteriori pdf.

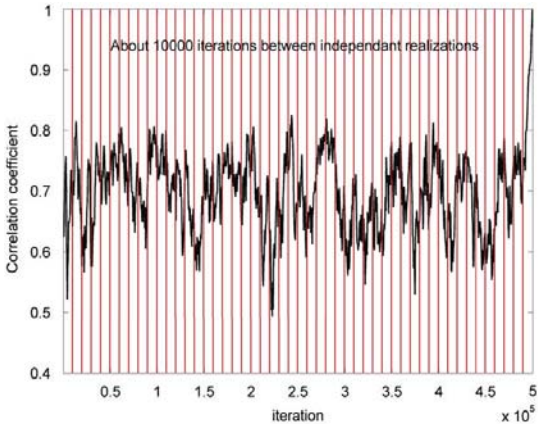


Fig. 6. Correlation coefficient between the last accepted model from the a posteriori pdf, and all other realizations of the a posteriori pdf.

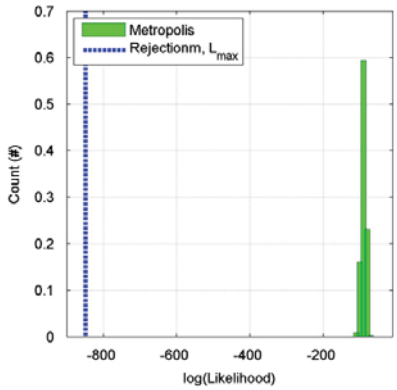


Fig. 7. Distribution of log-likelihood of the models considered in 500,000 iterations of the Metropolis sampler (green), and the one model of 500,000 considered model using rejection sampling with maximum-likelihood (blue dashed line). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Metropolis algorithm as considered above. The log-likelihood distribution of a posteriori accepted models is in the interval -105 to -75. However, the blue line indicates the maximum log-likelihood of -824 obtained after generating 500,000 independent realizations of the a priori pdf and evaluating the corresponding log-likelihood as part of running the rejection sampler. Thus, the 'best' model found after 500,000 realizations is very far from leading to a data fit within data uncertainties. Even if L_{max} could somehow be chosen around -68 (as indicated by the log-likelihood values of the accepted a posteriori models obtained from Metropolis sampling) the probability of locating just one realization from the a posteriori pdf using independent sampling of the a priori pdf, will be extremely low. The main problem with the rejection sampler is that it is computationally very inefficient for anything but very low dimensional problems. In general we suggest to make use of the extended Metropolis sampler to sample the a posteriori pdf of non-linear non-Gaussian inverse problems.

Sampling the a posteriori pdf using least-squares. As discussed in Hansen et al. (this issue), if the forward problem is linear, and a linear forward mapping operator given as `forward.G` is provided, then the a posteriori pdf can be sampled using least squares, kriging through error simulation or direct sequential simulation. Here we will consider using classical least squares type inversion, using `lsq_type='lsq'`. We will use exactly the same specification of the a priori model and the data model as used above.

To solve the linear Gaussian inverse problem using least squares type inversion, using the `ray`, `fat`, and `born` type forward model approximation, we use

```
forward.linear=1;
forward.type='ray';
forward.freq=10;
lsq_type='lsq';
nr=15;
```

```
% 'ray' type forward model
forward.type='ray';
[m_reals_ray,m_est_ray,Cm_est_ray] =
  sippi_least_squares(data,prior,forward,nr,lsq_type);
```

```
% 'fat' type forward model
forward.type='fat';
forward.freq=10;
[m_reals_fat,m_est_fat,Cm_est_fat] =
  sippi_least_squares(data,prior,forward,nr,lsq_type);
```

```
% 'born' type forward model
forward.type='born';
[m_reals_born,m_est_born,Cm_est_born] =
  sippi_least_squares(data,prior,forward,nr,lsq_type);
```

It is difficult to see any large difference between realizations from the a posteriori pdf using the three different types of forward models. Therefore, Fig. 8 shows the three a posteriori mean models, considering the (a) `ray`, (b) `fat`, and (c) `born` type forward model, which demonstrates that on average there is a difference between the solutions obtain with these different forward choices.

3.3.2. 2D non-linear inversion - AM24

We now consider the 2D data recorded between borehole AM2 and AM4, perpendicular to the data set recorded between borehole AM1 and AM3. We make the same assumptions about the a priori and the forward model as considered in the application of the extended Metropolis sampler above

```
D=load('AM24_data.mat');
forward.sources=D.S;
forward.receivers=D.R;
forward.type='eikonal';
```

As above we make use of the extended Metropolis algorithm to sample the a posteriori pdf. Fig. 9 shows 20 realizations of the 1D velocity from the a posteriori pdf considering the data sets AM13 and AM24, at location $x=2.5$ m, where the two profiles cross each other. Also shown is the mean of 200 a posterior realization for both data sets.

Fig. 9 reveals that where the two profiles intersect, the inferred velocity profile is quite similar even when the two data sets are inverted independently. In the top part of the model, where the consistency between realizations are weakest, the relative position of the relatively high velocity layers at depths of 2.8 m and 5 m is in agreement, while the velocity estimates of the more

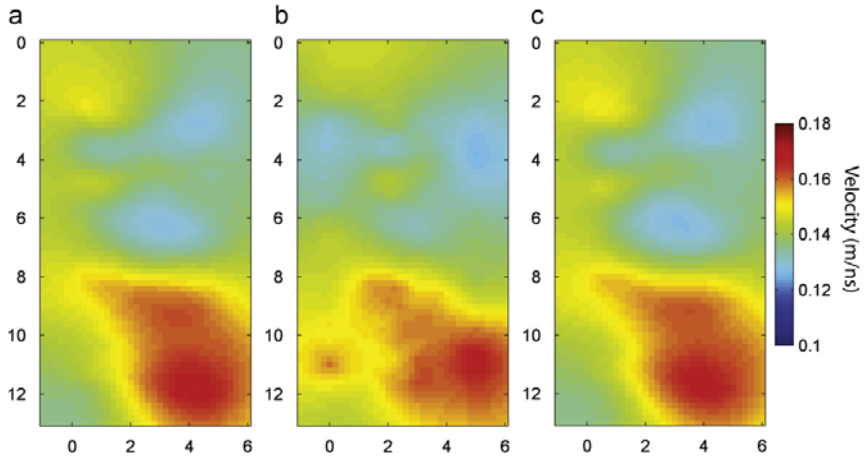


Fig. 8. Five realizations of the a posteriori pdf, using the (a) ray, (b) fat, and (c) Born type linear forward models.

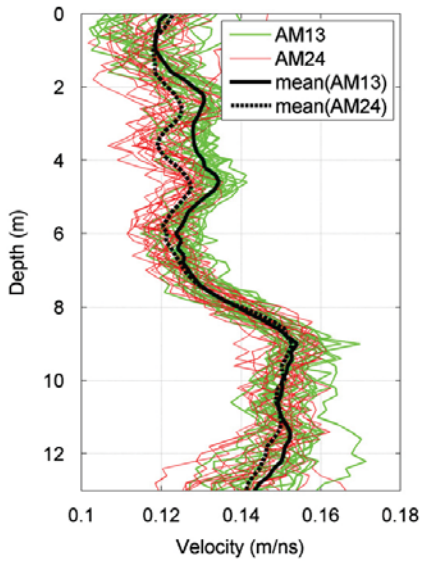


Fig. 9. Twenty realizations at $x=2.5$ considering data sets AM13 (green lines) and AM24 (red lines). The solid black line and dashed line show the corresponding average 1D velocity profile of 200 realizations of the a posteriori pdf. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

shallow parts differ only slightly. The reason for the observed inconsistencies can be related to the use of a 2D forward model describing data collected in a 3D world.

3.3.3. 3D inversion using a Gaussian a priori model—AM1234

Setting up an inversion using 3D data and a 3D parametrization of the a priori model is very similar to the 2D example above. Using the AM1234 data sets one can use

```
D=load('AM1234_data.mat');
forward.sources=D.S;
forward.receivers=D.R;
forward.type='eikonal';
```

The a priori model is identical to the one used above, except that a 3D parametrization needs to be specified. We also make use of a larger pixel size in order to keep the running time reasonable.

```
prior{im}.x=[-1 : .5 : 6];
prior{im}.y=[-1 : .5 : 6];
prior{im}.z=[0 : .5 : 13];
```

Sampling the a priori and a posteriori pdf, can be performed in exactly the same manner as done for the 2D cases above. Fig. 10 shows five independent realizations of the a posteriori pdf, obtained after the Metropolis algorithm has reached burn-in. Fig. 11 compare the mean of an a posteriori sample obtained from inverting the AM13, AM24, AM1234 data sets, at the location where the two 2D profiles intersect. Also shown is realizations from the a posteriori pdf corresponding to the AM1234 data set. Above 8 m depth the a posteriori mean is very similar for all cases. Below 8 m depth, the inferred velocity is higher inverting the 3D data set compared to the 2D data set.

3.3.4. 2D inversion with unknown covariance model properties

Most all inversion methods relying on a Gaussian a priori model, require that the properties of the covariance model, such as the mean, range, anisotropy, and variance are known prior to inversion. The choice of a priori covariance model highly affect the inversion result and, therefore, some work has been done to estimate a (prior) covariance model consistent with observed data, Asli et al. (2000), Hansen et al. (2008a), and Looms et al. (2010). As mentioned in Hansen et al. (this issue) and Le Ravalec et al. (2000) the FFTMA method allows for separating such structural properties of the covariance model from the random component. SIPPI allows such properties to act as model parameters, that can be inferred as part of an inversion.

To demonstrate this we use the same data and setup as used previously from the 2D travel time data set obtained between borehole AM1 and AM3, i.e. data set AM13, but where the a priori model is changed to allow for inference of the horizontal and vertical range.

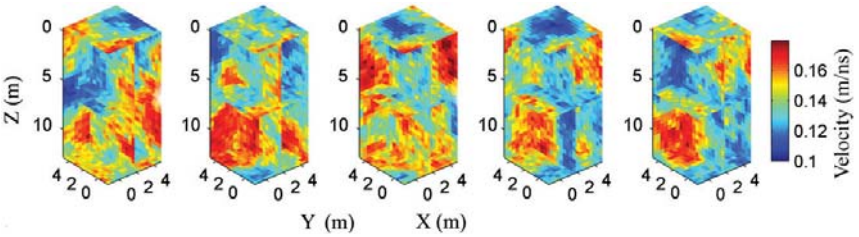


Fig. 10. Five realizations from the a posteriori pdf using the AM1234 3D data set.

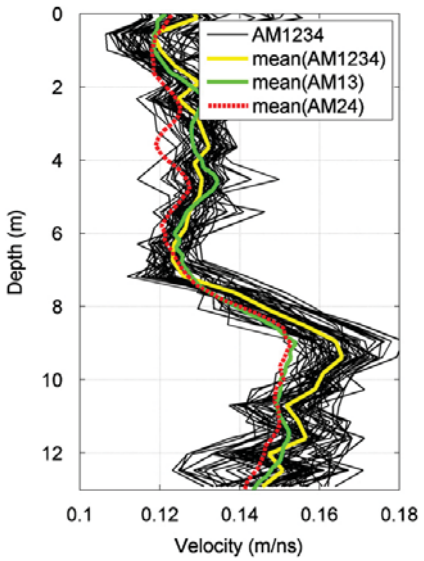


Fig. 11. 20 realizations of the a posterior pdf considering the 3D AM1234 data set of the center of the 3D grid where the two 2D profiles intersect (thin black) lines. Also shown is the mean of all a posteriori realizations considering the AM13 (green), AM24 (red), and AM1234 data sets (yellow). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

```
im=0;
% prior - HORIZONTAL RANGE
im=im+1;
prior{im}.type='gaussian';
prior{im}.m0=8 ;
prior{im}.std=6;
prior{im}.name='range_1' ;
prior{im}.prior_master=3;
prior{im}.norm=20;

% prior - VERTICAL RANGE
im=im+1;
prior{im}=prior{im-1};
prior{im}.name='range_2';


% prior - 2D VELOCITY FIELD
im=im+1;
prior{im}.type='FFTMA';
prior{im}.name='Velocity(m/ns)';
prior{im}.m0=0.145;
prior{im}.Va='.0003Sph(6)';
prior{im}.x=[-1 : .2 : 6];
prior{im}.y=[0 : .2 : 13];
```

Note that the only difference to the first example of inverting the AM31 data set with a known a priori covariance model, is the definition of two a priori parameters, named `range_1` and `range_2`. Also, these two prior structures point to the third prior structure (the `FFTMA` type prior) as their 'master', indicating which prior structure it belongs to. This ensures that when the value of such a prior model is updated, so is the value of covariance model of the corresponding `prior_master` structure.

A sample of this a priori model is shown in Fig. 12a. It is apparent that allowing variability in the ranges, determines an a priori model with much more a priori variability as compared to when the ranges is kept constant.

We now make use of the extended Metropolis sampler to sample the a posteriori pdf, in three cases where we use only 35 (every 20th observed data), 140 (every 5th observed data) and all 702 observed data, respectively. The subset of the data is chosen using the `data{id}.i_use=20 : 20 : 702` and `data{id}.i_use=5 : 5 : 702`, respectively. The corresponding samples from the a posteriori pdf is shown in Fig. 12b–d.

Because the horizontal and vertical ranges of the a priori covariance is also model parameters, the a posteriori distribution of these model parameters can also be quantified. Fig. 13 shows the 1D marginal posterior distribution of the horizontal and vertical range, respectively, using every (a) 20th, (b) 5th , and (c) all available observed data. When few observed data are used only very little information can be inferred about the ranges (red lines). But, as the number of data increases, so does the resolution of the range parameters. When all 702 data are used the 1D marginal a posteriori distributions of the ranges reveal that the horizontal range is relative long, between 7 m and 15 m, while the vertical range is better resolved with values between 4.8 m and 7 m. These findings are consistent with the result reported by Looms et al. (2010). Looms et al. (2010) find the range estimates priori to inversion of the travel time data, while in the present approach information about the ranges is inferred as part of the inversion.

As the number of considered observed data increase so does the resolution, which is seen as the differences between the a posteriori realizations become smaller. Thus increasing the amount of data leads to a better constrained posterior sample. It is, however, important to notice that the posterior statistics inferred from an a posteriori sample using only a subset of the data is consistent with the full solution: features that appear well resolved from a sample of the a posteriori pdf obtained using a subset of the data, will be consistent with the full inverse

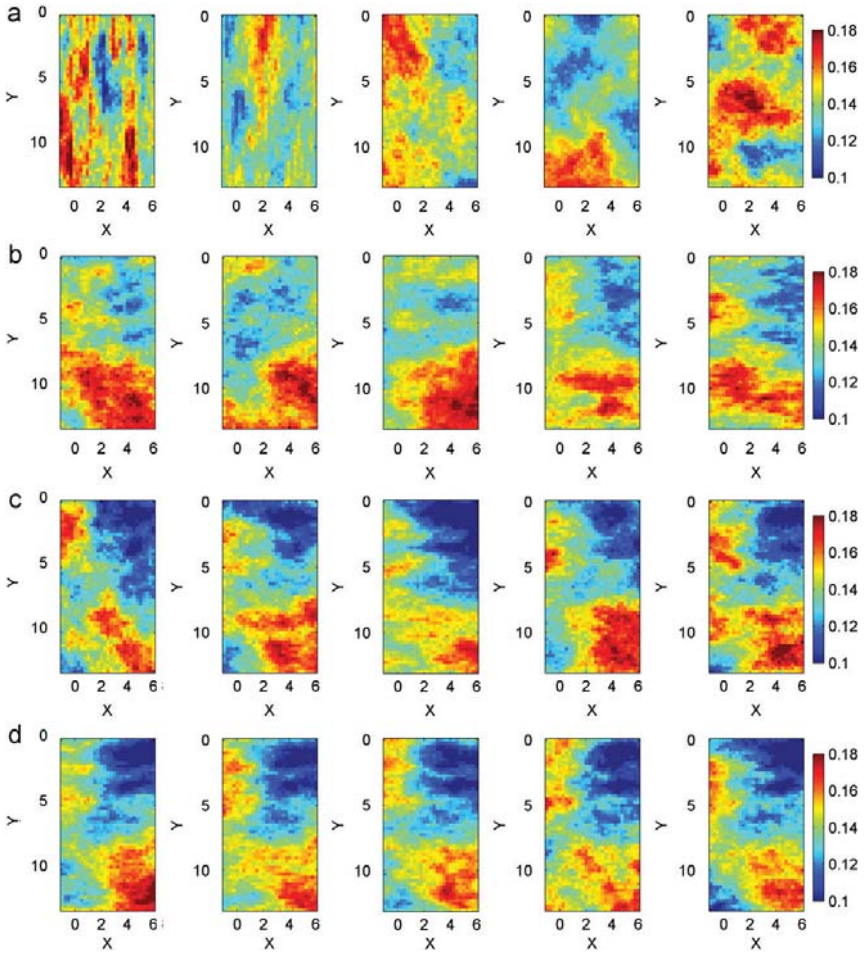


Fig. 12. Five realizations from the (a) a priori distribution and a posteriori distribution of the velocity field, using (b) 35, (c) 140, and (d) 702 observed data, respectively.

problem, unless some unaccounted for bias is present in data. There might be cases where the resolution provided by subset of the available data will be adequate. This will off course also result in an easier, more computationally efficient, sampling problem.

Traditional applications in inverse problems with Gaussian a priori models, rely on the existence of, or choice of, an a priori covariance model to describe spatial variability. The combination of the FFTMA prior model with the extended Metropolis sampler as implemented in SIPPI opens up new possibilities for solving non-linear inverse problems with unknown properties of the structural covariance model describing spatial variability.

3.3.5. 2D inversion with training image based prior

The a priori knowledge about the subsurface at Arrenæs does not readily call for a multiple point based a priori model, nor is such a model readily available. To demonstrate the use of a multiple point based a priori model, we generate a synthetic data set based on an a priori model defined by the training image in

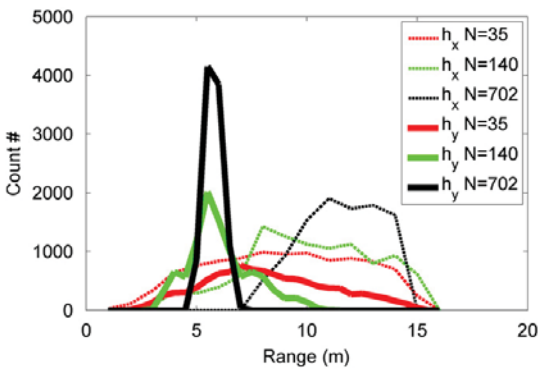


Fig. 13. 1D marginal a posteriori distribution of the horizontal (h_x) and vertical (h_x) range, using 35, 140 and 702 data observations, respectively.

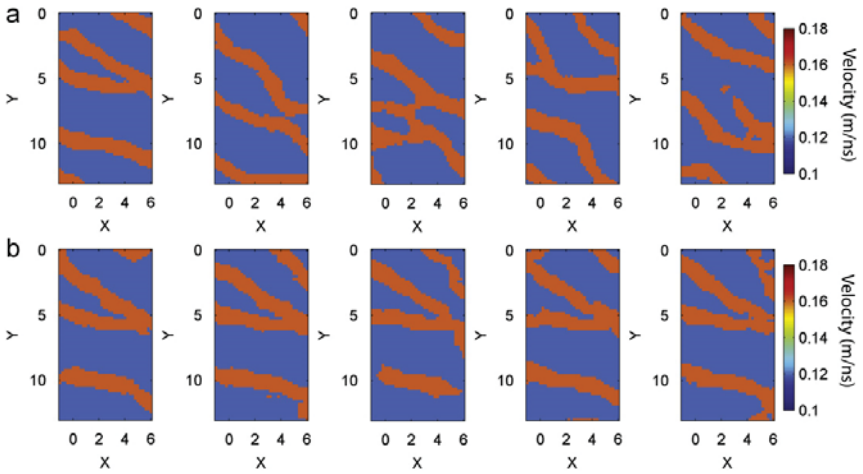


Fig. 14. Sample from the (a) a priori and (b) a posteriori distribution, considering the SNESIM type prior model, and synthetic data. The reference true model is the first of the five a priori realizations.

Fig. 4 in Hansen et al. (this issue), and the SNESIM type a priori model, Strebelle (2002), defined using

```
im=1;
prior{im}.type='SNESIM';
prior{im}.ti='snesim_std.ti';
prior{im}.index_values=[01] ;% optional
prior{im}.m_values=[.10.18] ;% optional
prior{im}.scaling=.75 ;% optional
prior{im}.rotation=30 ;% optional
```

Fig. 14a shows five realizations of this a priori model. The first model is chosen as the reference velocity model, from which synthetic data are computed by solving the forward problem. Finally some random Gaussian noise, according to the observed data uncertainties, are added to obtain an 'observed' data set.

```
id=1;
m_ref=sippi_prior(prior);
d_ref=sippi_forward(m_ref,forward,prior,data);
data{id}.d_obs=d_ref{1}+
  randn(size(d_ref{1})).*data{id}.d_std;
data{id}.Ct=0;
```

Then the Metropolis algorithm is run in the exact same manner as in the previous examples. Fig. 14b shows five realizations from the a posteriori pdf obtained by running the extended Metropolis algorithm.

This small example demonstrates that the difficulty of using a more complex a priori model using SIPPI, lies mostly in the difficulty to locate or choose such a model. Implementation wise there is only very little difference in choosing a simple covariance based prior model as opposed to a more complex prior model based on multiple point statistics.

4. Conclusions

We have demonstrated the use of the SIPPI toolbox to sample the solution to cross hole travel time tomographic inverse problems. A number of different forward models ranging from simple ray theory, based on high frequency wave-theory, to fat ray forward models based on finite frequency theory are available. We have demonstrated how such a tomographic inverse problem can be solved by sampling the a posteriori pdf, for a non-linear formulation of the inverse problem using the extended Metropolis algorithm for both 2D and 3D cases. We have also shown how least squares based techniques can be used to directly generate samples of the a posteriori pdf in the case of linear inverse Gaussian problems. Examples are based on a cross hole georadar data set. We have demonstrated that SIPPI facilitates a novel approach, based on a combination of the FFTMA method and the extended Metropolis sampler, that allow sampling the a posteriori pdf of linear and non-linear inverse problem with a Gaussian a priori model, where the properties of the covariance can be treated as parameters, and thus inferred as part of the inversion. Thus, the structural parameters defining the Gaussian a priori model, need not be known prior to inversion. All code and data is available using open licenses.

Acknowledgment

We thank DONG for financial support. SIPPI source code and the data from Arrenæs can be downloaded from http://sippi.sourceforge.net/.

References

Asli, M., Marcotte, D., Chouteau, M., 2000. Direct inversion of gravity data by cokriging. In: Kleingeld, W., Krige, D. (Eds.), Proceedings of the 6th International Geostatistics Congress, Cape Town, South Africa, April 10–14, pp. 64–73.

Buursink, M., Johnson, T., Routh, P., Knoll, M., 2008. Crosshole radar velocity tomography with finite-frequency Fresnel volume sensitivities. Geophysical Journal International 172 (1), 1–17.

Cordua, K.S., Nielsen, L., Looms, M.C., Hansen, T.M., Binley, A., 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 (1), 71–84.

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

Dafflon, B., Irving, J., Barrash, W., 2011. Inversion of multiple intersecting high-resolution crosshole GPR profiles for hydrological characterization at

the Boise Hydrogeophysical Research Site. Journal of Applied Geophysics 73 (4), 305–314.
Dahlen, F., Hung, S., Nolet, G., 2000. Fréchet kernels for finite-frequency travel-times-I. Theory. Geophysical Journal International 141 (1), 157–174.
Giroux, B., Gloaguen, E., 2012. Geostatistical traveltime tomography in elliptically anisotropic media. Geophysical Prospecting 60 (6), 1133–1149.
Gloaguen, E., Marcotte, D., Chouteau, M., 2005a. A non-linear tomographic inversion algorithm based on iterated cokrigriging and conditional simulations. In: Leuangthong, O., Deutsch, C. (Eds.), Geostatistics Banff 2004, vol. 1. Springer, pp. 409–418.
Gloaguen, E., Marcotte, D., Chouteau, M., Perroud, H., 2005b. Borehole radar velocity inversion using cokriging and cosimulation. Journal of Applied Geophysics 57 (4), 242–259.
Hansen, T.M., Cordua, K.S., Looms, M.C., Mosegaard, K. SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1, methodology. Computers & Geosciences, http://dx.doi.org/10.1016/j.cageo.2012.09.004, this issue.
Hansen, T.M., Looms, M.C., Nielsen, L., 2008a. Inferring the subsurface structural covariance model using cross-borehole ground penetrating radar tomography. Vadose Zone Journal 7 (1), 249.
Hansen, T.M., Mosegaard, K., Cordua, K.S., 2008b. Using geostatistics to describe complex a priori information for inverse problems. In: Ortiz, J.M., Emery, X. (Eds.), VIII International Geostatistics Congress, vol. 1. Mining Engineering Department, University of Chile, pp. 329–338.
Hansen, T.M., Journel, A.G., Tarantola, A., Mosegaard, K., 2006. Linear inverse Gaussian theory and geostatistics. Geophysics 71 (6), R101–R111.
Hansen, T.M., Mosegaard, K., 2008. VISIM: sequential simulation for linear inverse problems. Computers & Geosciences 34 (1), 53–76.
Hassouna, M., Farag, A., 2007. Multistencils fast marching methods: a highly accurate solution to the eikonal equation on cartesian domains. IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (9), 1563–1574.
Husen, S., Kissling, E., 2001. Local earthquake tomography between rays and waves: fat ray tomography. Physics of the earth and Planetary Interiors 123 (2), 127–147.
Irving, J., Knight, R., Holliger, K., et al., 2009. Estimation of the lateral correlation structure of subsurface water content from surface-based ground-penetrating radar reflection images. Water Resource Research 45 (12), W12404.
Irving, J., Knoll, M., Knight, R., 2007. Improving crosshole radar velocity tomograms: a new approach to incorporating high-angle traveltime data. Geophysics 72, J31.
Irving, J., Singha, K., 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. Water Resources Research 46.
Jensen, J., Jacobsen, B., Christensen-Dalsgaard, J., 2000. Sensitivity kernels for time–distance inversion. Solar Physics 192 (1–2), 231–239.
Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press.
Le Ravalec, M., Noetinger, B., Hu, L.Y., 2000. The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulations. Mathematical Geology 32 (6), 701-723.
Liu, Y., Dong, L., Wang, Y., Zhu, J., Ma, Z., 2009. Sensitivity kernels for seismic Fresnel volume tomography. Geophysics 74, U35–U46.
Looms, M.C., Hansen, T.M., Cordua, K.S., Nielsen, L., Jensen, K., Binley, A., 2010. Geostatistical inference using crosshole ground-penetrating radar. Geophysics 75 (6), J29.
Marquering, H., Dahlen, F., Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana–doughnut paradox. Geophysical Journal International 137 (3), 805–815.
Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory International Geophysics Series, vol. 45. Academic Press, revised edition.
Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse Problems 14, 405.
Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research 100 (B7), 12431–12447.
Nielsen, L., Looms, M.C., Hansen, T.M., Cordua, K.S., Stemmerik, L., 2010. Estimation of Chalk Heterogeneity from Stochastic Modeling Conditioned by Crosshole GPR Traveltimes and Log Data. No. 15 in Geophysical Developments Series. Society of Exploration Geophysicists, pp. 379–396 (Chapter 23).
Scholer, M., Irving, J., Looms, M.C., Nielsen, L., Holliger, K., 2012. Bayesian Markov-chain-Monte-Carlo inversion of time-lapse crosshole ground-penetrating radar to characterize the vadose zone at the Arrenæs field site, denmark. Vadose Zone Journal, http://dx.doi.org/10.2136/vzj2011.0153.
Sethian, J., Popovici, A., 1999. 3-D traveltime computation using the fast marching method. Geophysics 64 (2), 516–523.
Spetzler, J., Snieder, R., 2004. The Fresnel volume and transmitted waves. Geophysics 69, 653.
Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34 (1), 1–20.
Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM.
Tarantola, A., Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion. Reviews of Geophysics and Space Physics 20 (2), 219–232.
Topp, G., Davis, J., Annan, A., et al., 1980. Electromagnetic determination of soil water content: measurements in coaxial transmission lines. Water Resource Research 16 (3), 574–582.