

Ph.D. Thesis

Solving inverse problems through a smooth formulation of multiple-point geostatistics

Yulia Melnikova

National Space Institute
Division of Mathematical and Computational Geoscience
Technical University of Denmark
Kongens Lyngby
DTU SPACE-PHD-2013-



Preface

This thesis was submitted at the Technical University of Denmark, National Space Institute in partial fulfillment of the PhD requirements. The work presented in this thesis was carried out from June 2010 to October 2013. This project has been supervised by Professor Klaus Mosegaard, Associate Professor Alexander Shapiro and Professor Erling Stenby.

This project has been funded by the Danish Council for Strategic Research (——) and DONG Energy, which are gratefully acknowledged.

Acknowledgements

I would like to express my deep appreciation to my supervisor Klaus Mosegaard for his guidance, positive attitude and great support.

I would like to thank all current and former members of the Division of Mathematical and Computational Geoscience for their collaboration, advice and encouragement. My special thanks are extended to the staff of Scientific Computing Section in DTU Compute, where this project was initiated, for welcoming me in the new working environment. I wish to thank the IT support team of DTU Compute for saving my computers, warm Christmas celebrations and chocolate supplies.

I would like to thank all members of the BDC Club for making life brighter everyday. Also I wish to thank DTU Dancing for teaching me salsa and bachata. I am very grateful to the band Nickelback for their inspiring song "When we stand together", which was boosting my productivity in programming.

Finally, I wish to thank my parents, my friends and my significant other for their love, support and encouragement throughout this study.

Summary

Solving inverse problems through a smooth formulation of multiple-point geostatistics

In oil and gas sector accurate reservoir description play a crucial role in problems associated with recovery of hydrocarbons, risk estimation and predicting reservoir performance. Knowledge on reservoir properties can be inferred from measurements typically made at the surface by solving corresponding inverse problems. However, noise in data, non-linear relationships and sparse observations impede creation of realistic reservoir models. Including complex a priori information on reservoir parameters facilitates the process of obtaining acceptable solutions. Such a priori knowledge may be inferred, for instance, from a conceptual geological model termed a *training image*.

The main motivation for this study was the challenge posed by *history matching*, an inverse problem aimed at estimating rock properties from production data. We addressed two main difficulties of the history matching problem: existence of multiple, most often geologically unfeasible, solutions and high computational cost of the forward simulation. The developed methodology resulted in a new method for solving inverse problems with training-image based a priori information, when the computational time matters.

Specifically, we have proposed a smooth formulation of training-image based priors, which was inspired by the Frequency Matching method developed by our group earlier. The proposed smooth generalization, that integrates data and multiple-point statistics in a probabilistic framework, allows us to find solution by use of gradient-based optimization. As the result, solutions to an inverse problem may be obtained efficiently by deterministic search. We have applied the proposed methodology to the problem of history matching.

Both the smooth formulation and the Frequency Matching method find the solution by maximizing its posterior probability. This is achieved by introducing a closed form expression for the a priori probability density. We have defined an expression for the training-image based prior by applying the theory of multinomial distributions. Its combination with the likelihood function results in the closed form expression for defining relative posterior probabilities of the solutions.

Finally, we applied the developed smooth formulation to the problem of seismic inversion. The proposed methodology allows us to invert seismic reflection data for rock properties, namely for porosity, by integrating rock physics model into inversion procedure. Errors associated with conversion from depth to time are handled with a novel mapping approach.

This thesis reviews the latest developments in the field of geoscientific inverse problems with a focus on the history matching problem. The work contains detailed explanation of our strategies including both theoretical motivation and practical aspects of implementation. Finally, it is complemented by six research papers submitted, reviewed and/or published in the period 2010 - 2013.

Contents

Preface	iii
Acknowledgements	v
Summary	vii
1 Introduction	1
1.1 Probabilistic reservoir modeling	1
1.2 The history matching problem	2
1.3 History matching as an underdetermined, non-linear inverse problem . .	3
1.3.1 Forward model	5
1.4 Challenges of the history matching problem	7
1.4.1 The call for complex a priori information	8
1.5 Methods for solving geoscientific inverse problems with focus on the his- tory matching problem	8
1.5.1 Monte Carlo simulation	9
1.5.2 Ensemble Kalman Filter	9
1.5.3 Machine Learning methods	10
1.5.4 Gradient-based methods	11
1.5.5 Tailoring algorithms to the problem at hand	13
1.6 Our directions and contributions	14
2 Integration of complex a priori information	17
2.1 Methods of multiple-point statistics	17
2.1.1 Optimization-based approaches	19
2.1.2 Directions in the research	19
2.2 The Frequency Matching method	20
2.2.1 Solving inverse problems with the FM method	21
	ix

3	The smooth formulation	23
3.1	Motivation	23
3.2	The smooth formulation at a glance	25
3.3	Misfit with a priori information $f^d(\mathbf{m}, \mathbf{TI})$	26
3.3.1	The pseudo-histogram	26
3.3.1.1	Pattern similarity function	29
3.3.2	Similarity function	30
3.4	Implementation	32
3.4.1	Computing derivatives	32
3.4.2	Logarithmic scaling	33
3.4.3	Choosing optimization technique	34
3.5	Generating prior realizations	36
3.6	Solving inverse problems	36
4	Computing prior probabilities	41
4.1	Relation to Natural Language Processing	42
4.2	Computing prior probabilities of a discrete image given a discrete training image	44
4.2.1	Prior probabilities in the continuous case	46
5	Inversion of seismic reflection data using the smooth formulation	47
5.1	Introduction	47
5.2	Methodology	49
5.2.1	Seismic forward model	49
5.2.1.1	Rock physics model	50
5.2.1.2	Reflectivity series	51
5.2.2	Seismic sensitiviities	54
5.3	Numerical example	55
6	Conclusion	63
	Bibliography	65
	Appendices	73
A	Methods of unconstrained optimization	75
B	Defining prior probabilities: connection with chi-square and multivariate Gaussian	79
C	Ricker Wavelet	83

D Paper 1	85
E Paper 2	107
F Paper 3	113
G Paper 4	119
H Paper 5	125
I Paper 6	161

Introduction

1.1 Probabilistic reservoir modeling

Tools for creating reliable reservoir models are of a key interest for oil industry: prediction and optimization of hydrocarbon recovery, planning additional wells, estimation of risks — all these tasks rely on reservoir description. Reservoir characterization is a complex process of using multiple sources of data, expertise and numerical methods in order to describe subsurface structures, location of fluids and properties of rocks. Reservoir characterization can be treated as a task of data mining, where data vary in space and time, or it can be analyzed through non-quantitative expert knowledge. Typical data used in reservoir characterization include seismic surveys, well logs, production data and geological maps.

Understanding that the data are contaminated with noise and that the description of physical processes is not perfect leads to the necessity of probabilistic modeling. In this approach the available information about the subsurface is accounted for in accordance with its uncertainty and then integrated into the posterior probability distribution of the model param-

eters. Probabilistic modeling and uncertainty quantification are becoming paramount in reservoir characterization and form the philosophy of this study as well.

Challenges of integrated reservoir modeling include insufficient information content of the measurements, their complex non-linear relationship with reservoir properties, and different sensitivities. Complex prior information integrated into the posterior probability density function can drastically decrease uncertainties on the reservoir parameters. Modern methods rely, for instance, on the concept of the training image — a conceptual representation of expected geology in the formation. Geological expertise, a source of non-quantitative information, is also required to be integrated into the probabilistic framework.

This work was motivated by the history matching problem — the task of integrating production data for recovering reservoir properties. In this thesis we demonstrate how the production data can be efficiently constrained by complex a priori information within an efficient probabilistic framework.

1.2 The history matching problem

An important part of reservoir characterization consists in obtaining reservoir models, simulated response from which match the observed production data, i.e. solving the history matching problem.

Modeling of the well response is a challenge on its own, since very accurate knowledge of physical processes happening in the reservoir is needed, and the associated computational load to model these effects is huge. Response from the wells depends in a complex way on a great number of physical parameters of the system: pressure, temperature, well control, geological features, rock properties, distribution of fluids, fluid chemistry and PVT properties, etc. Typically, when the history matching problem is solved,

1.3. History matching as an underdetermined, non-linear inverse problem

many of these parameters are assumed to be known, for instance, from seismic surveys, laboratory experiments. In reality, all of them are subject to uncertainty, and lack of knowledge of some of them may strongly influence our predictions.

Geological features and rock properties, some of the most uncertain parameters, play a key role in reservoir flow behavior. While seismic data are widely used for resolving large scale geological structures, impermeable barriers, oil-water contacts and even porosity, production data are useful for inferring knowledge on connectivity within the formation and, consequently, on permeability.

Naive treatment of the history matching problem as a model calibration task is not only tedious (citing a manager from a famous oil company: "History matching of the reservoir took three weeks by trained personnel."), but also economically risky since no uncertainties are taken into account. Looking at the history matching problem as an inverse problem instead, provides us with tools for consistent data integration and uncertainty estimation.

1.3 History matching as an underdetermined, non-linear inverse problem

Let \mathbf{m} denote the reservoir model parameters and \mathbf{d}^{obs} the observed data. In inverse problem theory, the forward problem consists in finding data response \mathbf{d} given model parameters \mathbf{m} and a possibly non-linear mapping operator $g(\mathbf{m})$:

$$\mathbf{d} = g(\mathbf{m}) \tag{1.1}$$

In the history matching problem, the non-linear operator $g(\mathbf{m})$ represents the system of differential equations whose solution is implemented as a reser-

voir simulator (Sec. 1.3.1)

The inverse problem is then defined as a task of finding \mathbf{m} given the observed data \mathbf{d}^{obs} . According to inverse problem theory (Tarantola, 2005), the solution of an inverse problem is defined by its a posteriori probability density function $\sigma(\mathbf{m})$. Given some prior assumptions on the model parameters as the probability density $\rho(\mathbf{m})$, the a posteriori probability can be expressed as:

$$\sigma(\mathbf{m}) = kL(\mathbf{m})\rho(\mathbf{m}) \quad (1.2)$$

The likelihood function $L(\mathbf{m})$ defines how well a proposed model reproduces the observed data.

The likelihood function can typically be expressed through a data misfit function $S(\mathbf{m})$ as $L(\mathbf{m}) = k \exp(-S(\mathbf{m}))$ (Mosegaard and Tarantola, 1995), where k is a normalization constant.

For instance, assuming Gaussian uncertainties in the data, one writes down the likelihood function as:

$$L(\mathbf{m}) = k \exp(-\|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{\mathbf{C}_D}^2) \quad (1.3)$$

where \mathbf{C}_D is the data covariance matrix.

The a priori information $\rho(\mathbf{m})$, by definition, represents our probabilistic beliefs on model parameters before any data were considered. Often Gaussianity is assumed when assigning prior probabilities. Reservoir properties, unfortunately, do not follow Gaussian distributions, therefore modeling of $\rho(\mathbf{m})$ is an interesting and important part of the history matching problem. A large part of this thesis is devoted to assessing and integration of non-Gaussian a priori information.

However, before diving into the world of probability distributions, we first provide the mathematical framework for the processes happening in an oil

1.3. History matching as an underdetermined, non-linear inverse problem

reservoir, define equations of the forward model, and discuss issues associated with them.

1.3.1 Forward model

Oil reservoirs are typically modeled as a discretized porous medium with flow processes defined by mass conservation and Darcy laws.

In the general case, for multiphase multicomponent flow, the following system of equations is constructed (Aziz et al., 2005):

$$\frac{\partial}{\partial t} \left(\phi \sum_i S_i \rho_i y_{c,i} \right) + \nabla \cdot \sum_i (\rho_i y_{c,i} u_i) + \sum_i \tilde{q}_{c,i} = 0 \quad (1.4a)$$

$$u_i = -k \frac{k_{ri}}{\mu_i} (\nabla p_i - \rho_i g \nabla D) \quad (1.4b)$$

Equation 1.4a is a conservation equation for component c . Equation 1.4b defines the Darcy velocity for phase i . Here ϕ is porosity, S_i is saturation of phase i , ρ_i is density of phase i , $y_{c,i}$ is the mass fraction of component c in phase i ; k is the absolute permeability, k_{ri} , μ_i and p_i are relative permeability, viscosity and pressure of phase i , D is the vertical depth (positive downwards) and g is the gravitational acceleration; $\tilde{q}_{c,i}$ stands for the well term.

The unknowns in these equations are $y_{c,i}$, S_i , p_i . Production response \mathbf{d} (Eq. 1.1) can be calculated from these values, using, for instance, Peaceman's model (Peaceman, 1977).

In general, this system is computationally very demanding, and fast compositional solvers are in need. However, very often the number of components is assumed equal to the number of phases, and then the system of equations 1.4 turns into the black-oil model, the most popular model in reservoir simulation, as well as in reservoir characterization. In addition, depending on

1. INTRODUCTION

particular reservoir conditions, gravitational and compressibility effects can be neglected.

Commercial reservoir simulators, for instance Eclipse (Schlumberger GeoQuest, 2009), conventionally solve the system of flow equations (1.4) by implicit discretization, however the faster but less stable IMPES approach (implicit pressure, explicit saturation) is also an option. Still the time needed to simulate the forward response is high. Average models used in industry may consist of hundreds of thousands grid blocks and may require many hours of computation.

A good alternative to conventional simulators are streamline-based simulators (see, for example, StreamSim). In streamline simulators the pressure equation is solved on a conventional grid, however, values of saturations are propagated along the streamlines (lines tangent to the velocity field). This technique allows pressure to be updated much less frequently, which saves computational resources. Streamline simulators approximate flow very accurately when the fluid behavior is mainly defined by the heterogeneity of the subsurface, but not by diffusion processes as in unconventional reservoirs.

Reservoir model parameters needed to predict reservoir behavior depend on the complexity of the problem at hand. However, the minimum set includes porosity, permeability, relative permeability curves, fluid composition, initial depths of oil-water and oil-gas fluid contacts. Parameters updated in history matching are typically restricted to permeability and porosity, while other are assumed to be known. However, recent studies (Chen and Oliver, 2010b) have shown that including more variables such as oil-water contact and parameters of relative permeability may improve reconstruction of the primary rock properties.

Reservoir parameters possess different sensitivities to the different types of data. Such, water cut is sensitive to the properties along the streamlines and pressure data are good for recovering properties within vicinity of the

wells. All these observations stimulated research in localization techniques (Arroyo-Negrete et al., 2008; Emerick and Reynolds, 2011).

1.4 Challenges of the history matching problem

Having this theoretical base let us look at the particular challenges associated with the history matching problem.

First of all, the history matching problem is an underdetermined inverse problem: the number of observations is much smaller than the number of model parameters, due to the limited number of wells.

The second problem is related to the low information content in the production data, or, in other words, to the small number of independent data. Synthetic studies (Oliver et al., 2001; Van Doren et al., 1426) show that, despite integrating thousands of observations, only small percent of the model parameters can be identified by the data.

These difficulties combined with non-linearity make production data constraints weak for recovering realistic reservoir properties. As a result, multiple, often geologically unfeasible solutions to the history matching problem exist. Clearly, solutions inconsistent with geology cannot be used for prediction or optimization of the reservoir behavior. Large uncertainty of the solutions makes it impossible to draw probabilistic conclusions about the reservoir properties.

All these issues require special treatment of the history matching problem, such as including additional, constraining data. For example, geological expertise, seismic or well log data are all capable of facilitating the process of solving the history matching problem. Therefore research is concentrated on using additional constraints, and the problem is naturally growing into a problem of integrated reservoir modeling. Challenges associated with

conditioning reservoir parameters to different types of data are huge. Here are a few of them:

- different data scales
- high computational load
- uncertainty quantification

1.4.1 The call for complex a priori information

The necessity of geological constraints when solving inverse problems in geoscience is widely accepted. Recent advances in multiple-point geostatistics have made it a widely-used tool in geoscience problems, where reproduction of subsurface structure is required. Traditionally, multiple-point statistics is captured from training images that represent geological expectations of the subsurface (Guardiano and Srivastava, 1993). A review of the main techniques will be given in Chapter 2.

The use of a complex a priori information has two advantages: it assures more realistic solutions, and it facilitates the process of solving the inverse problem since it drastically constrains the solution space.

1.5 Methods for solving geoscientific inverse problems with focus on the history matching problem

The methods discussed below come from inverse problem theory, stochastic simulation techniques, data mining, optimization theory, pattern recognition, and machine learning. Very often they combine several techniques in one. The tremendous creativity is to a large extent spawned by the challenges of the history matching problem discussed above.

We start with the most sound tool for solving inverse problems: Monte-Carlo simulation.

1.5. Methods for solving geoscientific inverse problems with focus on the history matching problem

1.5.1 Monte Carlo simulation

Monte Carlo techniques are known to be the safest tool when an a posteriori pdf with complex a priori information is to be characterized (Mosegaard, 1998). Mosegaard and Tarantola (1995) suggested a modification of the Metropolis-Hastings algorithm (Metropolis et al., 1953) that allowed sampling of the posterior, without knowing an explicit expression of the prior distribution. In principle, it allows incorporation of a priori information of any complexity given an algorithm that samples the prior. When the extended Monte-Carlo sampling is performed, one suggests a perturbation of the current model m_k to a new model m'_k , according to the prior. The step is accepted with a particular probability:

$$P_{acc} = \begin{cases} 1 & \text{if } L(m'_k) > L(m_k) \\ L(m'_k)/L(m_k) & \text{otherwise} \end{cases} \quad (1.5)$$

in which case $m_{k+1} = m'_k$. In case of rejection, the model remains unchanged: $m_{k+1} = m_k$. This technique allows sampling the a posteriori pdf, such that the density of sampled models is proportional to the a posteriori pdf. This is referred to as *importance sampling*. Many examples of applications of Monte-Carlo methods to geoscientific inverse problems can be found in the literature (Hansen et al., 2012, 2008; Cordua et al., 2012b).

The extended Metropolis is easy to implement, but it requires a large number of forward simulations, both in the burn-in phase and in the sampling phase. In case of flow forward simulations (Sec. 1.3.1) sampling techniques are usually prohibitive.

1.5.2 Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is a sequential data assimilation technique, first introduced by Evensen (1994) for oceanography community. EnKF and its variants is one of the most popular methods for solving the history matching problem in research and industrial applications (Bianco et al., 2007; Haugen et al., 2008; Emerick and Reynolds, 2013; Lorentzen et al., 2013). This technique allows assimilation of multiple types of data,

to update a large number of parameters and to estimate uncertainty. It is easy to implement, since it is a derivative-free method, and reservoir simulator is treated as a black-box. Thorough review may be found in works of Aanonsen et al. (2009) and Oliver and Chen (2011)

The idea of EnKF consists in updating a set of models in a sequential manner, when new observation points are available. Model covariance is derived from ensemble members. For realistic representation of the uncertainty, a large number of ensemble members is needed, and it increases the computational load. The other well-known issue of EnKF is that covariance estimation based on the ensemble members tends to be corrupted by spurious correlations which results in collapse of the ensemble variability. Different covariance localization techniques can be found in Chen and Oliver (2010a), Emerick and Reynolds (2011), and Arroyo-Negrete et al. (2008).

EnKF performs best when the model parameters follow a Gaussian distribution and are related linearly to the data. For reservoir characterization these are strong assumptions, and therefore solutions of EnKF lack geological realism. Much work is being done to integrate non-Gaussian prior information into the EnKF framework. For instance, Sarma and Chen (2009) applied kernel methods to formulation of EnKF, which allowed preserving multiple-point statistics in final solutions. Recently Lorentzen et al. (2012) suggested using level-set functions to transform facies models first, and then applying EnKF to them. This resulted in very convincing non-smooth model updates.

1.5.3 Machine Learning methods

Efficiency of machine learning paradigm inspired researchers to use them in reservoir characterization. Machine learning methods are often defined as data-driven, since the relationship between parameters is inferred from the observed data. Demyanov et al. (2008) applied support vector regression (extension of the support vector machines) for reproducing geologically realistic porosity and permeability fields. The data that were used for training

1.5. Methods for solving geoscientific inverse problems with focus on the history matching problem

consisted of hard data and seismic observations. The technique was able to recover geological structures despite the presence of a strong noise in the data. The authors extended the technique from using single kernel towards multiple kernels (Demyanov et al., 2012)

Sarma et al. (2008) applied Kernel PCA for differentiable parameterization of multiple-point statistics. The formulation allows combining it with gradient-based optimization for solving history matching, for instance. We, actually, pursue the same goal in Chapter 3. The authors suggested using Karhunen-Loève expansion that allows deriving the necessary parameterization from the covariance matrix of the random process that generates realizations. The covariance matrix was obtained empirically from training image realizations. If the realizations were Gaussian, linear PCA would be enough for the parameterization. However, in order to reproduce the spatial features inherited into the training image, the Kernel PCA was needed. Kernel PCA allows mapping the original model parameters into some possibly extremely high-dimensional space where linear PCA can be applied to capture non-linearity of the parameters. As a result, new realizations can be generated in the feature space. Then, after solving the pre-image problem, the parameters are mapped back to the original space. When this formulation is combined with the data misfit term, the authors can solve the history matching problem.

1.5.4 Gradient-based methods

Gradient-based optimization is especially preferable for computationally heavy problems, such as the history matching problem. When gradient-based optimization is used, history matching is typically formulated as a minimization problem:

$$O(\mathbf{m}) = \frac{1}{2} \|d^{obs} - g(\mathbf{m})\|_{C_D}^2 + \|\mathbf{m} - \mathbf{m}_{prior}\|_{C_M}^2 \quad (1.6)$$

This formulation assumes Gaussianity of the model parameters. In Chapter 3 we show how we can use gradient-based optimization together with

1. INTRODUCTION

multiple-point statistics, substituting the traditional term $\|\mathbf{m} - \mathbf{m}_{prior}\|_{C_M}^2$ with a non-Gaussian prior misfit.

The gradient-based methods use data sensitivities to guide the solution towards a decrease of the objective function in each iteration. In this work we consider only methods for unconstrained optimization. When constraints are applied, an additional computational load is needed to respect boundaries of the feasible region. Therefore, traditionally, some logarithmic scaling is applied to the parameters (Tarantola, 2005; Gao and Reynolds, 2006). Appendix A reviews the main methods of unconstrained optimization such as steepest-descent, Newton method, Levenberg-Marquardt, and quasi-Newton methods. For details refer to Nocedal and Wright (2006).

In general, the model parameters are updated with the following iterative scheme:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{p}_k \quad (1.7)$$

Here α_k is the step length and \mathbf{p}_k is a search direction, that, depending on the method, integrates the function gradient (first derivatives) and possibly also the function's Hessian (second derivatives), or its approximation.

Despite the chosen technique, the knowledge of $\nabla O(\mathbf{m}_k)$ at each iteration k is required:

$$\nabla O(\mathbf{m}_k) = -G_k^T C_D^{-1} (d^{obs} - g(\mathbf{m}_k)) + C_M^{-1} (\mathbf{m}_k - m_{prior}) \quad (1.8)$$

where G_k is the sensitivity of the data with respect to the model parameters (Jacobian).

Optimization techniques differ from each other mainly by how the search direction is computed. The Levenberg-Marquardt method is typically called the most efficient method for finding least-squares solutions. Nevertheless it requires explicit computation of G_k . In contrast, the quasi-Newton method needs only the matrix-vector product $-G_k^T C_D^{-1} (d^{obs} - g(\mathbf{m}_k))$, which can

1.5. Methods for solving geoscientific inverse problems with focus on the history matching problem

be computed efficiently using, for instance, adjoints (Oliver et al., 2008).

In principle, G_k can be estimated by a finite-differences approach, however it involves a large number of forward computations. In the history matching problem G_k is frequently calculated using the adjoint equations (see Oliver and Chen (2011) for review) or derived using streamlines (Vasco et al., 1999; Datta-Gupta et al., 2001)

Another technique that found its application is simultaneous perturbation stochastic approximation (SPSA) of the derivatives (Spall, 1992). It can be used at early iterations to rapidly decrease the objective function.

1.5.5 Tailoring algorithms to the problem at hand

Great work was performed by Gao and Reynolds (2006) to adjust the gradient-based quasi-newton LBFGS method towards the needs of the history matching problem. They proposed two techniques that prevented the algorithm from suggesting abnormally high or low values of the rock properties. The first consists in damping the production data at early iterations, by artificially increasing the data covariance matrix. The other is based on applying constraining logarithmic scaling to the parameters. The advantage of this scaling is the possibility to stay within the framework of unconstrained optimization, while the outcome parameters are fixed within a certain range.

Another technique to improve performance of the gradient methods is to provide a good starting guess by conditioning to the facies information in wells (Zhang et al., 2003) or, for instance, employing results of seismic inversion.

1.6 Our directions and contributions

First of all, we solve inverse problems in a probabilistic framework following the Bayesian philosophy. It means that we strive to find a solution in the form of a probability distribution that reflect combinations of the likelihood function and a priori information.

Inverse problems in geoscience are often severely underdetermined and require use of complex a priori information on subsurface properties. Such information may be inferred, for instance, from training images by means of multiple-point statistics techniques. In Chapter 2 we discuss methods for integrating training-image based prior information into inverse problems. Appendix D reveals a recently suggested Frequency Matching method, that integrates and quantifies a priori information by means of multiple-point statistics. It is followed by a paper where the Frequency Matching method is integrated with the history matching problem (Appendix E).

Admitting computational challenges, we agree that a complete inference of the posterior may not be possible. Therefore we aim at finding an efficient way for solving the history matching problem, but staying within the Bayesian framework. Our approach consists in finding an ensemble of models that explain the data and honor the complex a priori information using gradient-based optimization.

The methodology is based on a smooth formulation of the multiple point statistics which is described in Chapter 3. The papers showing the development of this technique can be found in Appendices F, G. Appendix H contains a journal paper that discusses the method in greater details and summarizes its theoretical contributions. The proposed smooth formulation enables us to detect models belonging to different regions of high posterior probability, which can be further explored by sampling techniques.

In this thesis we propose a closed form expression for the a priori probability density function, that combined with the likelihood value, allows us to rank

solutions to the inverse problems in accordance with their relative posterior probabilities. The approach uses theory of multinomial distributions applied to the training image and its realizations. We provide motivation for it in Chapter 4 and show a computational example in Appendix H.

The proposed method can be applied to any inverse problem with a priori information defined by training images. In Chapter 5 we demonstrate inversion of 3D seismic reflection data combined with rock physics and multiple point geostatistics. Difficulties associated with conversion of reflection coefficients from depth-to-time are resolved by a novel mapping technique. Inversion of seismic data, that are very sensitive to the contrasts in physical properties of rocks, are able to provide a strong starting guess for the history matching problem. In addition, the developed method serves as a base for the joint inversion strategy.

Integration of complex a priori information

2.1 Methods of multiple-point statistics

Algorithms of multiple-point statistics (MPS) aim at integrating complex heterogeneity into models. Traditionally, they require a learning example, called a *training image* by Guardiano and Srivastava (1993), that is assumed to represent spatial phenomena. A training image may originate from outcrop modeling, geological expertise, and/or previous experience. Training images can represent both categorical (for example, rock types) and continuous properties (for example, porosity) of the system. In probabilistic modeling, training images can be used as a source of prior information on model parameters.

While the idea of capturing multiple-point statistics from training images belongs to Guardiano and Srivastava (1993), the successful story of MPS algorithms started with SNESIM, a pixel-based simulation technique developed by Strebelle (2002). Following some random path, the values of the

nodes are simulated by drawing from a distribution that is conditioned to the previously simulated nodes and hard data within a chosen neighborhood (defined by a template). The authors suggested using a tree structure to compute conditional probabilities efficiently. Nevertheless, SNESIM is memory demanding, and not all template sizes and number of categories can be afforded. Methods to improve its performance were developed (Straubhaar et al., 2011; Huang et al., 2013) SNESIM is one of the popular methods that is used today for simulating spatial properties.

An alternative approach for simulating spatial heterogeneity is to simulate several pixels at once: so-called pattern-based techniques. For instance, SIMPAT (Arpat and Caers, 2007) proceeds in a sequential manner, replacing group of pixels (defined by a template) with a pattern that is the most similar to the observed pixels. It implies scanning through the pattern database obtained from the training image every time and makes it computationally heavy. SIMPAT uses a concept of similarity between the patterns which is an important concept in our work as well (Chapter 3).

Another pattern-based method called FILTERSIM (Zhang et al., 2006) uses linear filters to project patterns into a lower-dimensional space and derives all similarities in a faster manner. This approach also allows simulating realizations of continuous training images.

Further improvement of pattern-based simulation happened when DISTPAT (Honarkhah, 2010, 2011) was developed. This method uses multidimensional scaling (MDS) to investigate pattern variability in a low-dimension space.

Direct sampling (Mariethoz et al., 2010a) is essentially a pixel-based method, also using a concept of similarity between the patterns. It is efficient, since no conditional probabilities are needed to be computed. Instead, the pixel value is taken directly from a pattern that is the most similar to already simulated nodes surrounding this pixel.

2.1.1 Optimization-based approaches

Another family of the methods solves some optimization problem, at each iteration estimating how well the multiple-point statistics is honored by the model. One recent example of such methods is the Frequency Matching method (Lange et al., 2012a), see details in Sec.2.2.

The method developed in this thesis (Chapter 3) also, in general, allows to simulate spatial features learned from the categorical training image. In this iterative, smooth formulation of MPS pixels are perturbed all at once. The result of such simulation will be an image, which pixel values are almost discrete as required by the training image. In rigorous sense, this result can not be called realization of the prior. Nevertheless, statistically speaking, its properties are close to the pure discrete realizations. The details are discussed in (Chapter 3).

2.1.2 Directions in the research

Development of the methods for MPS is an active topic in reservoir characterization. One can formulate the following research problems associated with the training images:

- source of realistic training images (Sech et al., 2009)
- use of non-stationary training images(De Vries et al., 2009)
- use of continuous training images (Zhang et al., 2006; Mariethoz et al., 2010b)
- improving pattern reproduction (Cordua et al., 2012a)
- improving speed of simulation (Mariethoz et al., 2010b; Huang et al., 2013)
- finding balance between variability of patterns and quality of their reproduction (Tan et al., 2013)

- accounting for uncertainty in training images (Jafarpour and Khodabakhshi, 2011)

2.2 The Frequency Matching method

The Frequency Matching method (Lange et al. (2012a), Appendix D) is a novel method for solving the inverse problems using multiple-point geostatistics of the training image. The core of the Frequency Matching method is to represent and compare images using the frequency distributions of their patterns. Consider the training image and the 2×2 search template applied to it (Fig. 3.2a). The corresponding frequency distribution of the patterns is shown in Fig. 2.1b.

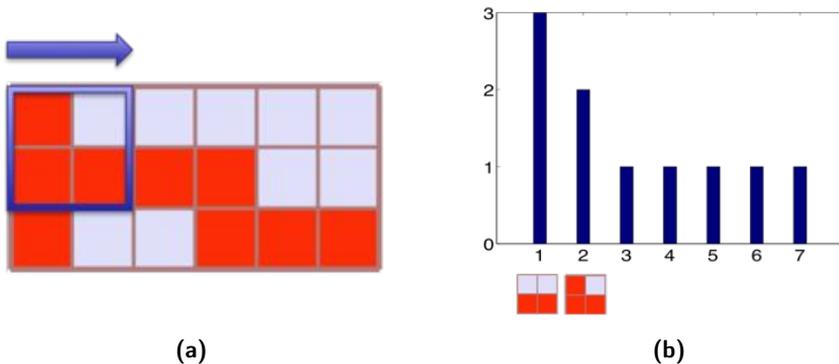


Figure 2.1: Example of a binary discrete image (a) and its frequency distribution for a 2×2 template (b)

The FM method suggests a way to generate models statistically similar to the training image by minimizing the distance between histograms of the training image and the model. This is equivalent to the approach of SNESIM, only here conditional probabilities are computed from the his-

togram (see Lange et al. (2012a)).

When the histogram misfit is combined with the data misfit term, an optimization problem is solved for finding the maximum a posteriori solution. Lange et al. (2012b) find the maximum a posteriori solution of the inverse problem by minimizing the following sum of misfits:

$$\mathbf{m}^{MAP} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_D}^2 + \alpha f(\mathbf{m}, \mathbf{TI}) \right\} \quad (2.1)$$

2.2.1 Solving inverse problems with the FM method

Lange et al. (2012a) used the FM method for reconstructing seismic velocities of rocks in a crosshole tomography setup. The solution fitted the data and honored the multiple point statistics. In addition, due to the rich ray coverage, the solution strongly resembled the true model.

The Frequency Matching method was applied to the history matching problem for a small 3D synthetic model (Melnikova et al. (2012), Appendix E) . In its original formulation, the FM optimization was implemented through simulated annealing, which prevented the algorithm from being stuck in local minima for the objective function. For the history matching problem, this approach was too expensive, therefore we used a simple rule (a 'greedy search') where a proposed model was accepted only if the value of the objective function (Eq. 2.1) decreased. In order to improve the performance of the method, fast proxy models based of the streamline simulations were used. The results demonstrated successful reproduction of spatial features, nevertheless the production data turned out to be weak constraints.

The smooth formulation

In this chapter we discuss a novel method for representing multiple point statistics through a smooth formulation.

3.1 Motivation

The smooth formulation of the training-image based prior was motivated by one of the history matching problem challenges: the high cost of the forward simulation. The Frequency Matching method (Lange et al. (2012a), Chapter 2) suggests an optimization framework, where model parameters are perturbed in such a way that fit to the data and consistency with the a priori information are iteratively improving. The resulting solution explains the data and honors multiple point statistics inferred from the training image. The formulation of the FM method requires the model parameters to take discrete values. It implies solving a combinatorial optimization problem (Eq. 2.1), which results in a large number of forward simulations needed to achieve the match. As it was demonstrated in Melnikova et al. (2012) (Appendix E), the number of forward simulations for a modest 2D history

3. THE SMOOTH FORMULATION

matching problem was of the order of tens of thousands.

The smooth formulation derived from the simple ambition of moving fast towards a high posterior region. One way of doing that is to suggest a differentiable parameterization of multiple-point statistics, that, at the end, would allow a gradient-based search for the solution. Traditionally MPS methods operate with categorical images, given the training image is categorical. The smooth formulation, in contrast, allows model parameters to take continuous values, although prohibited by the prior, while moving towards regions of high prior and posterior probabilities.

Consider Figure 3.1: When one is trapped in a maze, the ability to move is limited by the walls of the maze. This is analogous to the way the formulation of the FM method is limited by the discrete model space. One's dream would be to have the power of going through the maze walls to achieve the goal. This is the intuitive idea behind the smooth formulation.

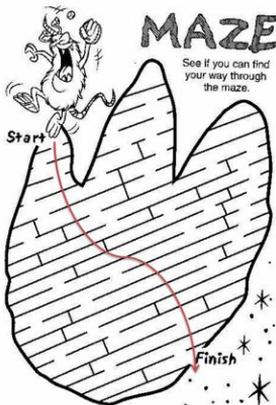


Figure 3.1: A parallel between a maze and a discrete model space. The red line depicts a direct way to the goal through prohibited states.

3.2 The smooth formulation at a glance

The smooth formulation allows us to gradually change a possibly continuous starting model \mathbf{m} into a model $\mathbf{m}^{\text{HighPosterior}}$ of high posterior probability, i.e. into one that honors both data and multiple-point statistics of the training image \mathbf{TI} .

The following differentiable objective function is the mainstay of the method:

$$O(m) = \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_D}^2 + f^d(\mathbf{m}, \mathbf{TI}) \quad (3.1)$$

where the first term is the data misfit term, and $f^d(\mathbf{m}, \mathbf{TI})$ is the misfit between the pattern histogram of \mathbf{TI} and a generalized pattern histogram for \mathbf{m} , defined for non-categorical images. Superscript d is used to emphasize that $f^d(\mathbf{m}, \mathbf{TI})$ is a differentiable function of the pixel values of \mathbf{m} . This distinguishes this formulation from the original formulation of the Frequency Matching method (Eq. 2.1)

Expression 3.1 can be minimized efficiently by gradient-based optimization techniques, since the proposed expression $f^d(\mathbf{m}, \mathbf{TI})$ for integration of multiple-point statistics is analytically differentiable.

The suggested formulation has several advantages:

- the solution is obtained fast due to the gradient-based search
- the solution honors the complex a priori information
- the optimization can be initiated from any convenient starting guess

The following explanation proceeds as follows. In Section 3.3 we show how $f^d(\mathbf{m}, \mathbf{TI})$ is constructed. In Section 3.4 we discuss some important practical aspects of the implementation. In Section 3.5 we show how by minimizing $f^d(\mathbf{m}, \mathbf{TI})$ the training-image based prior can be honored. In Section 3.6 we formulate a workflow for solving inverse problems by means of the smooth formulation. We apply the workflow to a synthetic history matching problem.

3.3 Misfit with a priori information $f^d(\mathbf{m}, \mathbf{TI})$

The value of $f^d(\mathbf{m}, \mathbf{TI})$ should reflect how well the multiple-point statistics of the discrete training image \mathbf{TI} is represented in a possibly continuous image \mathbf{m} . Recall that the Frequency Matching method uses the chi-square distance between frequency distributions of the training image and the model to evaluate the reproduction of MPS. Clearly, no frequency distribution for a continuous image can be constructed.

Instead, we propose to operate through a pseudo-histogram concept, that, similarly to the frequency distribution, estimates proportions of patterns, but can be computed for any continuous image.

3.3.1 The pseudo-histogram

Our notation is presented in Table 3.1. Notice the notation **image**, which implies that all symbols from Table 3.1 containing **image** as a superscript are defined both for the model and the training images.

The pseudo-histogram $H^{d,\mathbf{image}}$, where $\mathbf{image} \in \{\mathbf{m}, \mathbf{TI}\}$, similarly to the frequency distribution, reflects pattern statistics. It has two additional properties:

- it is differentiable with respect to the model parameters \mathbf{m}

3.3. Misfit with a priori information $f^d(\mathbf{m}, \mathbf{TI})$

Table 3.1: Notation

Notation	Description
\mathbf{TI}	training image, categorical
\mathbf{m}	model (test image), can contain continuous values
$\mathbf{image} \in \{\mathbf{m}, \mathbf{TI}\}$	image (training or test)
\mathbf{T}	scanning template
$H^{d,\mathbf{image}}$	pseudo-histogram of \mathbf{image}
$N^{\mathbf{TI},un}$	number of unique patterns found in \mathbf{TI}
$N^{\mathbf{image}}$	number of patterns in \mathbf{image}
$pat_j^{\mathbf{TI},un}$	j^{th} unique pattern in \mathbf{TI}
$pat_i^{\mathbf{image}}$	i^{th} pattern of pixels from \mathbf{image} .

- it can be computed for any continuous and discrete image

In order to construct the pseudo-histogram $H^{d,\mathbf{image}}$ the following steps are to be performed:

First, we scan through the \mathbf{TI} with the template \mathbf{T} and save its unique (not repeating) patterns as a database with $N^{\mathbf{TI},un}$ entries.

For $\mathbf{image} \in \{\mathbf{m}, \mathbf{TI}\}$ the pseudo-histogram $H^{d,\mathbf{image}}$ is defined as a vector of the length equal to the number of unique patterns in the \mathbf{TI} . Unique patterns of the training image define categories of the discrete patterns, whose proportions need to be matched during the optimization.

Our approach is based on the following idea: a continuous pattern $pat_i^{\mathbf{image}}$ does not fit to a single discrete pattern category, instead it contributes to all $N^{\mathbf{TI},un}$ categories.

3. THE SMOOTH FORMULATION

Consequently, the j^{th} element $H_j^{d,\mathbf{image}}$ reflects the “contribution” of all patterns found in \mathbf{image} to $pat_j^{\mathbf{TI},un}$:

$$H_j^{d,\mathbf{image}} = \sum_{i=1}^{N_{\mathbf{image}}} p_{ij}^{\text{sim}} \quad (3.2)$$

where p_{ij}^{sim} defines the level of similarity between $pat_i^{\mathbf{image}}$ and $pat_j^{\mathbf{TI},un}$.

We define p_{ij}^{sim} such that it equals 1 when $pat_i^{\mathbf{image}}$ is pixel-wise equal to $pat_j^{\mathbf{TI},un}$. We choose p_{ij}^{sim} to be based on the Euclidean distance between pixel values of the corresponding patterns:

$$p_{ij}^{\text{sim}} = \frac{1}{(1 + A t_{ij}^k)^s} \quad (3.3)$$

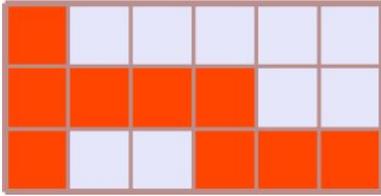
where $t_{ij} = \|pat_i^{\mathbf{image}} - pat_j^{\mathbf{TI},un}\|_2$ and A, k, s are user-defined parameters.

Notice the following property:

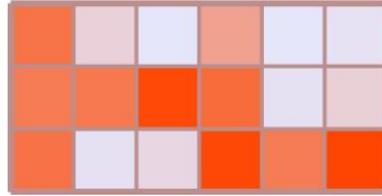
$$p_{ij}^{\text{sim}} = \begin{cases} 1 & t_{ij} = 0 \\ \in (0, 1) & t_{ij} \neq 0 \end{cases} \quad (3.4)$$

The properties of the pattern similarity function (Eq. 3.3) are discussed in Sec. 3.3.1.1

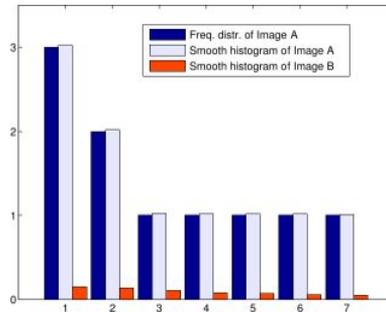
The pseudo-histogram computed for the discrete Image A (Fig 3.2a) is shown in Fig. 3.2c by light-blue color, compare it with the frequency distribution (dark-blue). Figure 3.2b shows a continuous image, while in Fig. 3.2c one can see its pseudo-histogram depicted by the orange color. Notice the small counts everywhere: indeed, according to Eq. 3.4, this image does not contain patterns sufficiently close to those observed in the training image.



(a) Discrete Image A



(b) Continuous image B



(c) Frequency distribution of Image A; Smooth histogram of Image A; Smooth histogram of Image B; 2x2 template applied

Figure 3.2: Frequency distribution and its approximation

3.3.1.1 Pattern similarity function

The choice of A , k , s in Eq. 3.3 is very important: on one side, they define how well the pseudo-histogram approximates the true frequency distribution; on the other side, they are responsible for “smoothing” and, consequently, for the convergence properties. Figure 3.3 reflects how different values of k , s with fixed $A = 100$ influence the shape of the pattern sim-

ilarity function defined by Eq.3.3 (distance is normalized). Our empirical conclusion is that values $A = 100$, $k = 2$, $s = 2$ are optimal. Compare them (Fig 3.3) to the extreme case $A = 100$, $k = 1$, $s = 2$ where the majority of patterns have a close-to-zero contribution. These parameters are applicable after t_{ij} has been normalized by the quantity representing maximum possible Euclidean distance between the discrete patterns.

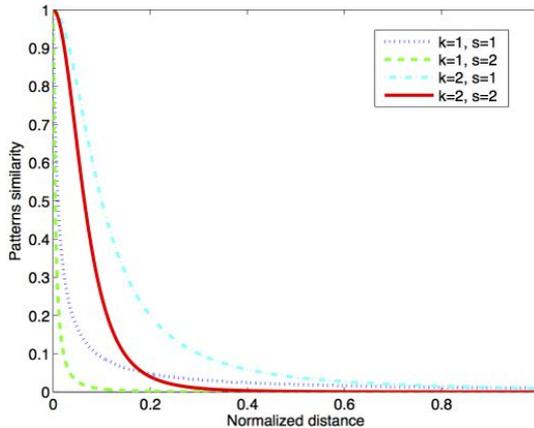


Figure 3.3: Patterns similarity function

3.3.2 Similarity function

Per definition, statistically similar images will have similar pseudo-histograms. Therefore we introduce the similarity function:

$$f^d(\mathbf{m}, \mathbf{TI}) = \frac{1}{2} \sum_{i=1}^{N^{\mathbf{TI}, un}} \frac{(H_i^{d, \mathbf{m}} - H_i^{d, \mathbf{TI}})^2}{H_i^{d, \mathbf{TI}}} \quad (3.5)$$

Essentially, it is a weighted L2 norm, where the role of the weight parameter is played by the smooth histogram of the training image. The suggested measure favors patterns that are encountered less frequently in the training image and facilitates proper reproduction of the training image features. If

3.3. Misfit with a priori information $f^d(\mathbf{m}, \mathbf{TI})$

the number of patterns in the training image $N^{\mathbf{TI}}$ differs from the number of patterns in the model $N^{\mathbf{m}}$, we multiply $H_i^{d, \mathbf{TI}}$ by the following ratio:

$$r = \frac{N^{\mathbf{m}}}{N^{\mathbf{TI}}} \quad (3.6)$$

The choice of the similarity function is validated in Chapter 4, where the expression of prior probability density function is explicitly derived.

3.4 Implementation

3.4.1 Computing derivatives

Methods of gradient-based optimization typically require a procedure that evaluates first-derivatives of the objective function. In this section we show how to analytically compute the gradient of the expression $f^d(\mathbf{m}, \mathbf{TI})$ (which is part of Eq. 3.1).

By definition:

$$\nabla f^d(\mathbf{m}, \mathbf{TI}) = \left[\frac{\partial f^d}{\partial m_1}, \dots, \frac{\partial f^d}{\partial m_n} \right]^T \quad (3.7)$$

From Eq. 3.5 it reads:

$$\nabla f^d(\mathbf{m}, \mathbf{TI}) = \begin{bmatrix} \frac{\partial H_1^{d,\mathbf{m}}}{\partial m_1} & \frac{\partial H_2^{d,\mathbf{m}}}{\partial m_1} & \dots & \frac{\partial H_{N^{un}}^{d,\mathbf{m}}}{\partial m_1} \\ \frac{\partial H_1^{d,\mathbf{m}}}{\partial m_2} & \frac{\partial H_2^{d,\mathbf{m}}}{\partial m_2} & \dots & \frac{\partial H_{N^{un}}^{d,\mathbf{m}}}{\partial m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_1^{d,\mathbf{m}}}{\partial m_n} & \frac{\partial H_2^{d,\mathbf{m}}}{\partial m_n} & \dots & \frac{\partial H_{N^{un}}^{d,\mathbf{m}}}{\partial m_n} \end{bmatrix} \begin{bmatrix} \frac{H_1^{d,\mathbf{m}} - H_1^{d,\mathbf{TI}}}{H_1^{d,\mathbf{TI}}} \\ \frac{H_2^{d,\mathbf{m}} - H_2^{d,\mathbf{TI}}}{H_2^{d,\mathbf{TI}}} \\ \vdots \\ \frac{H_{N^{un}}^{d,\mathbf{m}} - H_{N^{un}}^{d,\mathbf{TI}}}{H_{N^{un}}^{d,\mathbf{TI}}} \end{bmatrix} \quad (3.8)$$

As it was defined in Sec. 3.3.1, $H_j^{d,m}$ reflects contribution of all patterns found in \mathbf{m} , therefore from Eq. 3.2:

$$\frac{\partial H_j}{\partial m_z} = \sum_{i=1}^{N^m} \frac{\partial p_{ij}^{sim}}{\partial m_z} = \sum_{i=1}^{N^m} -Aks(1 + At_{ij}^k)^{(-s-1)} t_{ij}^{k-1} \frac{\partial t_{ij}}{\partial m_z} \quad (3.9)$$

where $z = 1, \dots, n$.

Notice, that $\frac{\partial t_{ij}}{\partial m_z} = 0$ if $m_z \notin \text{pat}_i^{\mathbf{m}}$.

Otherwise, if $\text{pat}_i^{\mathbf{m}} = [v_{i,1} \cdots v_{i,N}]^T$, and $\text{pat}_j^{TI,un} = [u_{j,1} \cdots u_{j,N}]^T$, where N is the number of pixels in the pattern, we get:

$$t_{ij} = \|\text{pat}_i^{\mathbf{m}} - \text{pat}_j^{\mathbf{TI}}\|_2 = \sqrt{(v_{i,1} - u_{j,1})^2 + \cdots + (v_{i,N} - u_{j,N})^2} \quad (3.10)$$

And, therefore:

$$\frac{\partial t_{ij}}{\partial m_z} = \frac{v_{i,s} - u_{j,s}}{\|\text{pat}_i^{\mathbf{m}} - \text{pat}_j^{\mathbf{TI}}\|_2} \quad (3.11)$$

where $v_{i,s} = m_z$.

3.4.2 Logarithmic scaling

The model parameters in reservoir characterization typically take positive values (as, for instance, permeability), or are constrained to be in a certain range (as, e.g., porosity). However, an iterative process in unconstrained optimization may suggest a perturbation that will violate these boundaries. One way to stay within the efficient framework of unconstrained optimization is to rescale parameters. We suggest using the logarithmic scaling (Gao and Reynolds, 2006):

$$x_i = \log \left(\frac{m_i - m^{low}}{m^{up} - m_i} \right) \quad (3.12)$$

where $i = 1, \dots, n$, and 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, of the parameters. This log-transform does not allow extreme values of the model parameters. We choose $m^{low} < \min(\mathbf{TI})$ and $m^{up} > \max(\mathbf{TI})$.

Notice, that for practical reasons we apply the same log-transformation to the training image (denoted as \mathbf{TI}_{log}). Then we minimize $f^d(\mathbf{x}, \mathbf{TI}_{log})$, which is equivalent to minimizing the original $f^d(\mathbf{m}, \mathbf{TI})$.

3.4.3 Choosing optimization technique

Choice of optimization technique depends on the size of the problem, as well on availability of sensitivities of the data with respect to the parameters. The discussion on unconstrained optimization can be found in Appendix A.

For the history matching problem, quasi-Newton methods are recommended Oliver and Chen (2011). Among many gradient methods (steepest-descent, Newton, Levenberg-Marquardt) the family of quasi-Newton methods excel in the balance between efficiency and simplicity of implementation. These methods use information from the second derivatives, similar to the Newton method, however, instead of computing the Hessian directly, use its smart approximations. Quasi-Newton methods perform especially well on large scale problems.

The Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) is one of the most popular quasi-Newton techniques. The update is calculated as:

$$x_{k+1} = x_k + \alpha_k p_k \tag{3.13}$$

where α_k is the step length and p_k is the search direction. The search direction is defined as follows:

$$p_k = -B_k^{-1} \nabla f_k \tag{3.14}$$

where B_k is the approximation of the Hessian at the k 'th iteration. The BFGS formula dictates the iterative update:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{3.15}$$

where $s_k = x_{k+1} - x_k = \alpha_k p_k$ and $y_k = \nabla f_{k+1} - \nabla f_k$.

In this work we used its modified version called the limited memory BFGS, which is especially suitable for the large scale problems. The limited memory BFGS method does not require storing fully dense approximations of the Hessian, instead only few vectors are stored to implicitly represent the approximation.

3.5 Generating prior realizations

Figure 3.4 shows the workflow for generating prior realizations using the proposed smooth formulation of multiple-point statistics. Examples of generating prior realizations can be found in Melnikova et al. (2013) (Appendix H) Notice, that while some pixels contain intermediate values, statistical features of the training image as well as expected sharp contrasts of the features are successfully reproduced.

3.6 Solving inverse problems

Solving the optimization problem 3.1 directly may result in an unbalanced fit of the data and prior information. This may happen because, while the data misfit term is derived directly from the definition of the likelihood, the a priori information is taken into account approximately via the smooth formulation.

In order to provide a fair balance between two terms, we pursue the idea of scaling the terms, making them dimensionless. One of the easiest ways to combine objective functions into a single function is to use the weighted exponential sum (Marler and Arora, 2004). We put equal weights on two misfit terms and the exponent equal to 2. In addition, we apply the logarithmic scaling described in Sec. 3.4.2.

This leads to the final expression for the objective function:

$$O^*(\mathbf{x}) = \left(\frac{\frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m}(\mathbf{x}))\|_{C_d}^2 - u^*}{u^*} \right)^2 + \left(\frac{f^d(\mathbf{x}, \mathbf{TI}_{log}) - f^*}{f^*} \right)^2 \quad (3.16)$$

where u^* and f^* are the desired values of the misfits.

Figure 3.5 demonstrates the workflow of the proposed methodology.

The figures show the performance of the method: initial guesses of permeability field (Fig. 3.6a) are fed into the workflow (3.5) and constrained by the multiple-point statistics and production data. The intermediate results after 50 iterations are shown in Fig. (3.6b), and the final solutions achieved after 150 iterations are shown in Fig. (3.6c).

3. THE SMOOTH FORMULATION

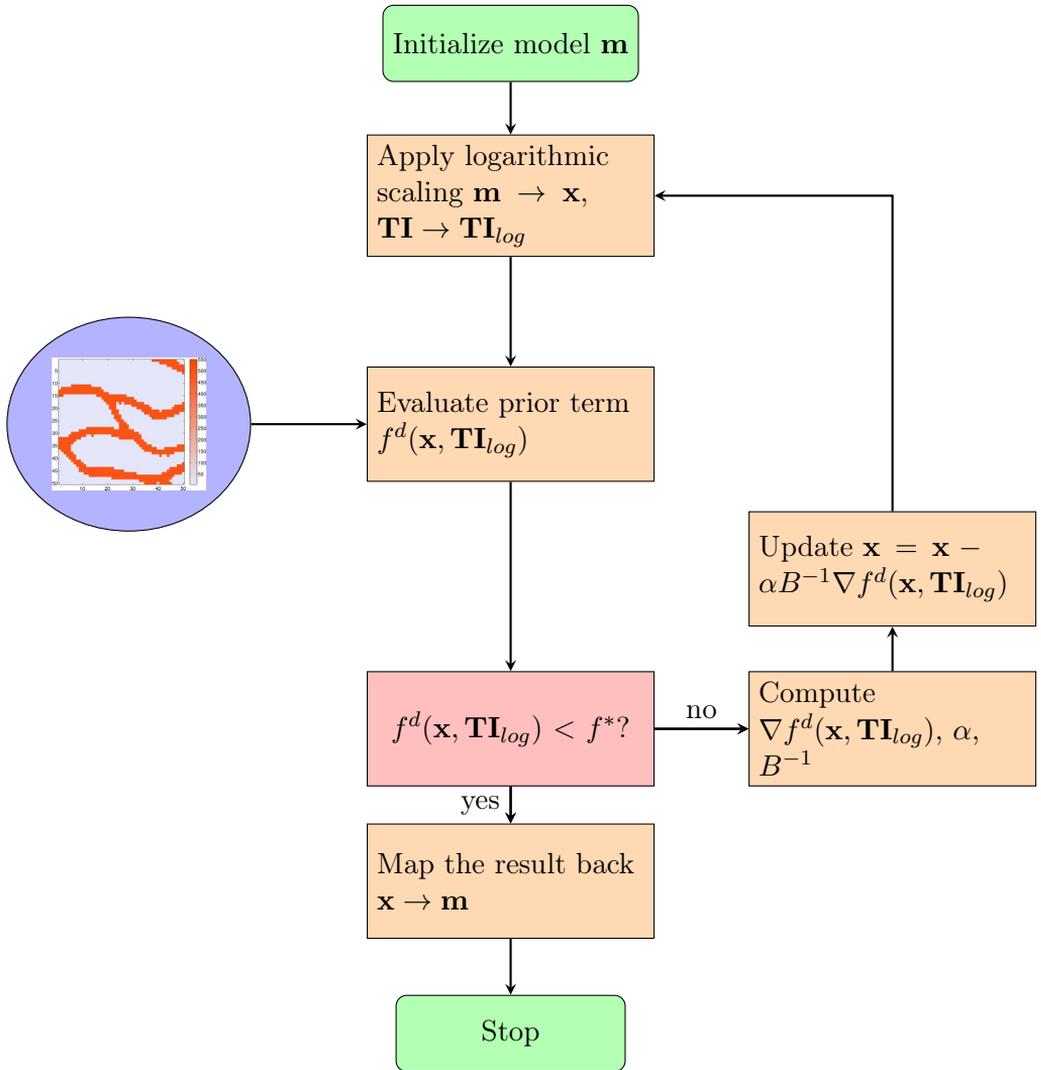


Figure 3.4: Flowchart

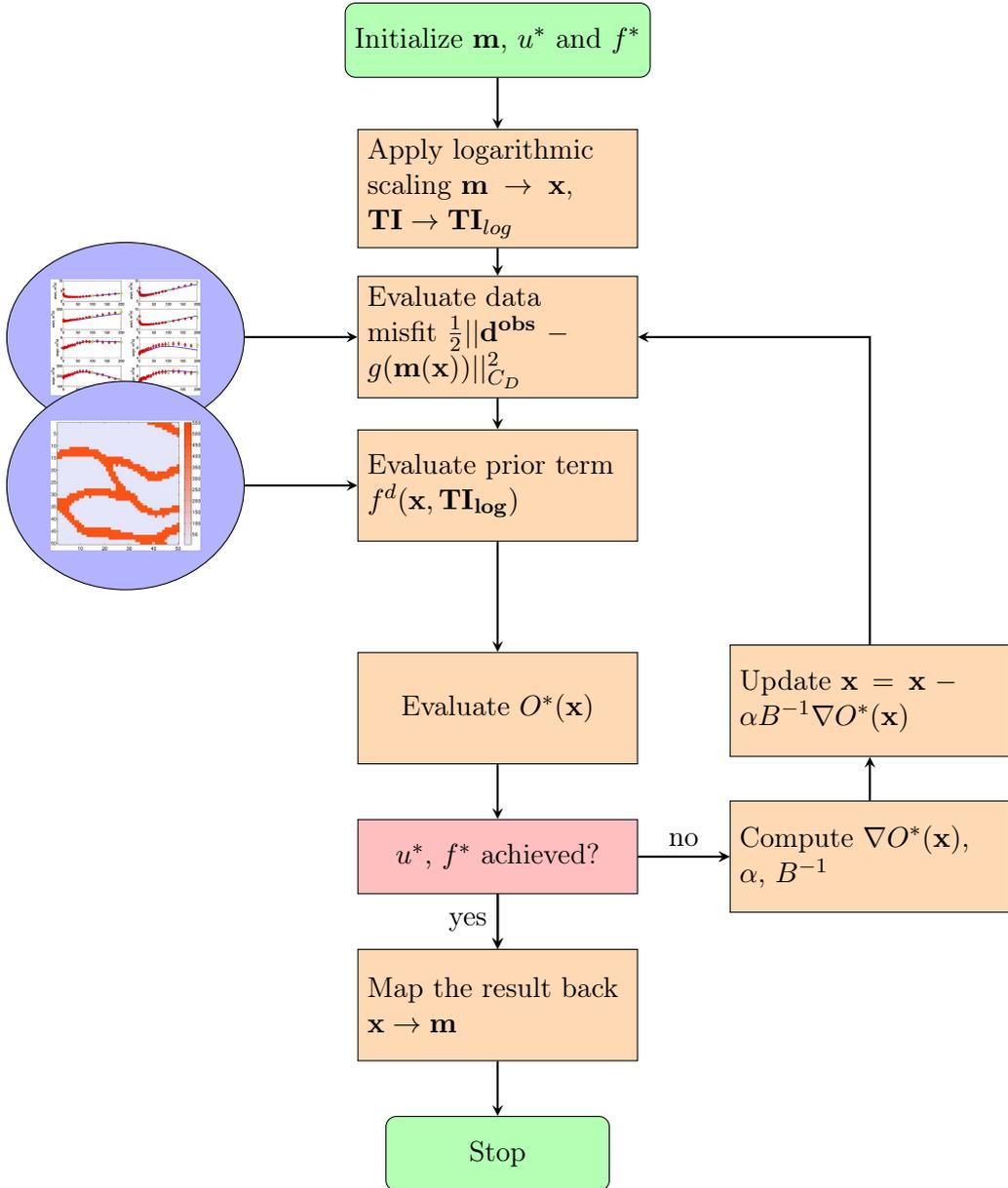


Figure 3.5: Flowchart

3. THE SMOOTH FORMULATION

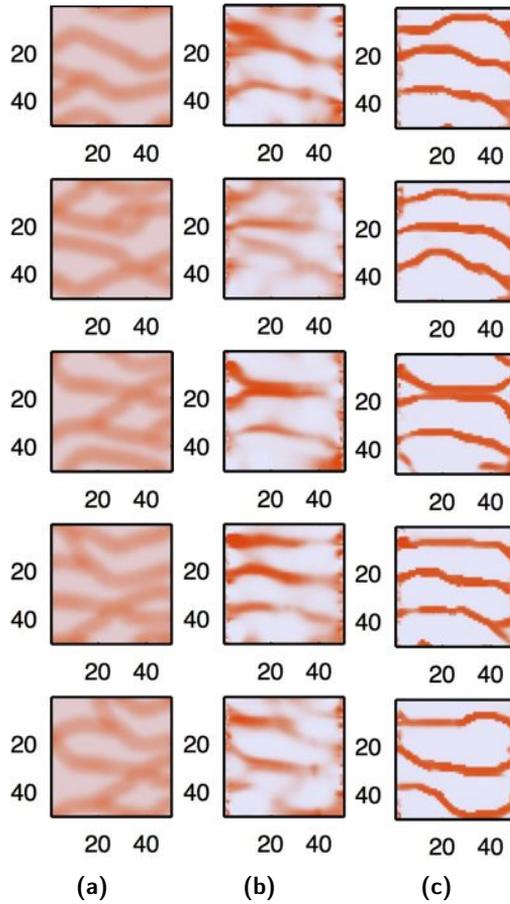


Figure 3.6: (a) Starting models, (b) Models after 50 iterations, (c) Models after 150 iterations

Computing prior probabilities

Methods for solving inverse problems that aim at maximizing the a posteriori probability through optimization (Lange et al., 2012b; Melnikova et al., 2013) require a closed form expression for the a priori probability density function to be known. Deriving such an expression for complex a priori information represented, for instance, by a training image, is not a trivial task.

Lange et al. (2012b) were first to suggest a closed form expression for the a priori PDF from a training image, based on its histogram of patterns. It was defined by means of the chi-square distance between pattern histograms, however it was lacking a definition of the normalization constant. Cordua et al. (2012b) suggested an alternative formulation of the a priori PDF, using the Dirichlet distribution. Both approaches assume an unknown theoretical distribution that generated the training image.

In Melnikova et al. (2013) we formulate an expression for the a priori probability density function assuming that the training image itself is capable of providing the necessary information on prior probabilities. This approach is more practical, since, indeed, the a priori knowledge is formed by the

observed training images. At first, however, let us compare the training image with the training dataset that is used in the field of Natural Language Processing (NLP) and show the common challenges.

4.1 Relation to Natural Language Processing

Researchers in Natural Language Processing also operate with prior probabilities. In such application as speech recognition, specific combinations of words, for instance sentences that 'make sense', are assigned non-zero prior probabilities. These probabilities are usually calculated from training datasets. Since the amount of data is usually insufficient, these probabilities can only be estimated approximately.

Consider a small example adopted from Marler and Arora (2004). Let us consider a small training set consisting of only three sentences: "JOHN READ MOBY DICK", "MARY READ A DIFFERENT BOOK", and "SHE READ A BOOK BY CHER".

Now, consider a test sentence "CHER READ A BOOK". The probability that the word "READ" follows the word "CHER" is defined as the number of counts of "CHER READ" divided by the number of occurrences word CHER followed by any word w (including end of string):

$$p(\text{READ}|\text{CHER}) = \frac{c(\text{CHER READ})}{\sum_w c(\text{CHER } w)} = \frac{0}{1} \quad (4.1)$$

From the given training set we conclude that the probability is zero. Assuming the bigram language model, where the probability of a word depends on the preceding word only (the Markov assumption), we obtain the following expression for the probability of our test sentence:

$$\begin{aligned} p(\text{CHER READ A BOOK}) &= p(\text{CHER}|\langle\text{bos}\rangle)p(\text{READ}|\text{CHER}) \\ & p(\text{A}|\text{READ})p(\text{BOOK}|\text{A})p(\langle\text{eos}\rangle|\text{BOOK}) = 0 \end{aligned} \quad (4.2)$$

where $\langle \text{bos} \rangle$ and $\langle \text{eos} \rangle$ mean 'beginning of string' and 'end of string', respectively.

Clearly, the probability of the test sentence is underestimated, since it is a meaningful combination of words with some probability to occur. The problem lies in the small size of the training data. Getting back to the speech recognition, one can ask "what is the probability of a string s given an acoustic signal A ?". Through the Bayesian rule, it can be found as:

$$p(s|A) = \frac{p(A|s)p(s)}{p(A)} \quad (4.3)$$

If $p(s)$, the prior probability of the sentence, was underestimated and was assumed to be zero, then the speech recognition algorithm fails, regardless the clarity of the acoustic signal. When we use a training image as a source of a priori information, we find ourselves in the exactly the same situation. Information obtained from the training image is not sufficient to assign correct prior probabilities.

The problem of insufficient training data can evidently be solved by considering bigger dataset. Another way to improve probability calculations is to apply *smoothing* techniques (Chen and Goodman, 1999). In general, these techniques aim at making distributions more uniform, increasing near-zero probabilities and decreasing high probabilities.

Marler and Arora (2004) review smoothing techniques and conclude that the Kneser-Ney smoothing is the most efficient approach. However, it is beyond the scope of this thesis to explore techniques for optimal smoothing of training-image-based priors. In our work we use a simpler technique called *absolute discounting*, that performs sufficiently well.

4.2 Computing prior probabilities of a discrete image given a discrete training image

Our idea consists in representing an image as an outcome of a multinomial experiment (see also Cordua et al. (2012a)). Consider two categorical images: training and test. Assume that a pattern in the test image is a multiple-point event that leads to the success for exactly one of K categories, where each category has a fixed probability of success p_i . By definition, each element H_i in the frequency distribution \mathbf{H} indicates the number of times the i^{th} category has appeared in N trials (the number of patterns observed in the test image). Then the vector $\mathbf{H} = (H_1, \dots, H_K)$ follows the multinomial distribution with parameters N and \mathbf{p} , where $\mathbf{p} = (p_1, \dots, p_K)$

$$P(\mathbf{H}) = P(H_1, \dots, H_K, N, p_1, \dots, p_K) = \frac{N!}{H_1! \dots H_K!} p_1^{H_1} \dots p_K^{H_K} \quad (4.4)$$

We assume that the vector of probabilities \mathbf{p} is inferred from the frequency distribution of the training image $\mathbf{H}^{\mathbf{T}\mathbf{I}}$: normalizing its entries on the total number of counts, we obtain the probabilities of success.

In general, the histogram of the training image is very sparse, therefore many categories of patterns will be assigned zero probabilities. It means that if a test image has a single pattern that is not encountered in the training image, its prior probability, as follows from Eq. 4.4, will be zero. It happens due to insufficient information in the finite training image; it is very likely that many of the unobserved patterns in the training image have some non-zero probabilities to be observed in a new situation.

Since there is no information about the probabilities of the patterns not encountered in the training image, we assume them to be equal to a small positive number ε . To make the sum of p_i equal to one, we subtract a small number γ from all non-zero bins of $\mathbf{H}^{\mathbf{T}\mathbf{I}}$:

$$p_i = \begin{cases} \frac{H_i^{\mathbf{T}\mathbf{I}} - \gamma}{N^{\mathbf{T}\mathbf{I}}} & H_i^{\mathbf{T}\mathbf{I}} > 0 \\ \varepsilon & H_i^{\mathbf{T}\mathbf{I}} = 0 \end{cases} \quad (4.5)$$

4.2. Computing prior probabilities of a discrete image given a discrete training image

where $\gamma = \varepsilon(K - N^{\mathbf{TI}, \text{unique}})N^{\mathbf{TI}}/N^{\mathbf{TI}, \text{unique}}$

After p_i having been defined, $P(\mathbf{H})$ can be computed through:

$$\log(P(\mathbf{H})) = \log\left(\frac{N!}{H_1! \cdots H_K!}\right) + \sum_{i=1}^K H_i \log(p_i) \quad (4.6)$$

We apply Stirling's approximation:

$$\log(n!) = n \log n - n + O(\log n) \quad (4.7)$$

Defining $I = \{i : H_i > 0\}$ we have:

$$\begin{aligned} \log\left(\frac{N!}{H_1! \cdots H_k!}\right) &= \log(N!) - \sum_{i \in I} \log(H_i!) \approx N \log N - N - \\ &\sum_{i \in I} (H_i \log(H_i) - H_i) = N \log N - \sum_{i \in I} H_i \log(H_i) \end{aligned} \quad (4.8)$$

And finally,

$$\log(P(\mathbf{H})) \approx N \log N + \sum_{i \in I} H_i \log\left(\frac{p_i}{H_i}\right) = \sum_{i \in I} H_i \log\left(\frac{N p_i}{H_i}\right) \quad (4.9)$$

Then

$$-\log(P(\mathbf{H})) \approx \sum_{i \in I} H_i \log\left(\frac{H_i}{N p_i}\right) \quad (4.10)$$

Substituting H_i with $N p_i + \varepsilon_i$ and applying a Taylor expansion of the second order one arrives to the chi-square distance divided by two:

$$-\log(P(\mathbf{H})) \approx \frac{1}{2} \sum_{i \in I} \frac{(H_i - N p_i)^2}{N p_i} \quad (4.11)$$

Further, if we denote $\mathbf{h} = \mathbf{H}/N$, Eq. 4.9 is transformed:

$$\log(P(\mathbf{H})) \approx \sum_{i \in I} N h_i \log\left(\frac{p_i}{h_i}\right) = - \sum_{i \in I} N h_i \log\left(\frac{h_i}{p_i}\right) = -N D_{KL}(h||p) \quad (4.12)$$

where $D_{KL}(h||p)$ is the Kullback-Leibler divergence, a dissimilarity measure between two probability distributions h and p . In other words, it defines the information lost when the theory (training image) is used to approximate the observations (test image).

Now, given a discrete image, one can compute its relative prior probability using Eq.4.9. Alternative way of computing prior probabilities consists in reformulating Eq.4.11 through the multivariate Gaussian (Appendix B). However it is less precise and was not applied in this work.

4.2.1 Prior probabilities in the continuous case

Consider a situation when the pixel values are continuous, but close to integer values, as inferred from the training image. One way to define its (approximate) prior probability is to round off the value, thereby obtaining a discrete image and use Eq. 4.9. However, this may influence the datafit, and consequently the likelihood and the posterior.

From a practical point of view, pixel values that differ a little do not spoil perception of spatial features and are sufficient for the end-user. For this reason, such patterns can be considered as a success in the multinomial experiment. Therefore, the above considerations are valid for near-integer models generated by our smooth approach.

In addition, notice that Eq.4.11 justifies our choice of similarity function (Eq. 3.5). Indeed, by minimizing expression 3.5 we minimize the value defined by Eq.4.11 as well. Examples of computing relative a priori probabilities can be found in Melnikova et al. (2013).

Inversion of seismic reflection data using the smooth formulation

This chapter describes the methodology for inverting seismic reflection data for rock properties using the smooth formulation.

5.1 Introduction

Seismic reflection data are widely used in reservoir characterization for resolving geological structure and properties. Seismic reflection data are especially attractive for inversion, as they are highly sensitive to the contrasts in the subsurface. Nevertheless there are several difficulties that prevent us from obtaining a unique solution: presence of noise in data, uncertainties associated with data processing, uncertain wavelet estimation, inaccurate rock-physics model and uncertainty in conversion from depth to time (Bosch

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

et al., 2010).

Commonly, seismic inversion aims at estimating elastic properties. Inversion for rock properties is a more complicated task, since the relationship between elastic parameters (impedances, velocities, elastic moduli) and rock properties (porosity, permeability) is often non-linear and uncertain. Typically this task is performed sequentially: first the elastic parameters are inverted and then the rock properties are estimated at a postprocessing stage.

We propose a new technique for inversion of seismic reflection data, that incorporates rock physics modeling and multiple-point geostatistics. The inverse problem is formulated as an optimization problem, where the a priori information is taken into account using the smooth formulation described in Chapter (3). The algorithm uses a gradient-based optimization in order to find a solution that belongs to the high posterior region, i.e. one that matches the data and multiple-point statistics of the training image representing spatial features in the subsurface.

We model the three dimensional seismic forward response by calculating zero-offset seismograms at each location at the surface using 1D convolution. This approach allows us to compute 3D seismic data almost at no cost. We use a simple rock physics model based on the Wyllie equation (Wyllie et al., 1956) that connects porosity, which is the model parameter we aim to invert for, with the acoustic impedance. In addition, we neglect fluid substitution effects. The method can be easily extended to a more complex rock physics model, for instance, to the one described in Avseth et al. (2000), since the elastic properties are still differentiable with respect to porosity.

This study is challenged by the difficulties of depth-to-time conversion. We propose a differentiable way for converting reflections from depth to time, that helps in reconstruction of geological features in correct locations.

An advantage of the proposed method is its efficiency when searching for

a solution with high posterior probability – solutions that can be further explored by sampling techniques. In addition, the method can be combined with history matching problem into a joint inversion strategy.

5.2 Methodology

We formulate the inverse problem in a probabilistic framework (Tarantola, 2005) where the solution is characterized by its a posteriori probability density function. Models of high posterior probability can then be obtained by maximizing the values of posterior PDF or minimizing the corresponding sum of misfits:

$$\mathbf{m}^{\text{HighPosterior}} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{d}_{\text{obs}}^{\text{seis}} - g^{\text{seis}}(\mathbf{m})\|_{C_D}^2 + f^d(\mathbf{m}, \mathbf{TI}) \right\} \quad (5.1)$$

Here $f^d(\mathbf{m}, \mathbf{TI})$ represents the misfit with multiple-point statistics of the training image as defined in Chapter 3. In the context of the current study, \mathbf{m} denotes porosity.

5.2.1 Seismic forward model

An important task in the development of the method is to propose a seismic operator g^{seis} , such that its Jacobian \mathbf{G}^{seis} with respect to porosity can be computed analytically.

We consider a discretized subsurface model consisting of $n_x \times n_y \times n_z$ gridblocks of size $\Delta x \times \Delta y \times \Delta z$ meters. We model the seismic response in the form of seismograms measured at each gridblock at the top layer of the model, so $n_x \times n_y$ seismograms are computed. The forward response is then defined as a vector consisting of concatenated seismograms. Each seismogram is computed with the same number of sampling points n_s , therefore size of the data vector is $n_s n_x n_y$.

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

The matrix of seismic response \mathbf{S} consisting of $n_x n_y$ rows, where each row represents a single seismogram, is computed using convolution of the reflectivity series \mathbf{R} with a wavelet, assumed to be known:

$$\mathbf{S} = w * \mathbf{R} = \mathbf{R}\mathbf{W} \quad (5.2)$$

where w is the discretized wavelet, and \mathbf{W} is the matrix operator performing convolution with w . In this study we used the Ricker wavelet, see Appendix C for details.

Reflection coefficients are modeled as a function of porosity differentiable everywhere. In this study we propose a differentiable conversion from depth to time, that helps in locating geological layers correctly. But first we introduce the rock physics model that provides us with a link to porosity.

5.2.1.1 Rock physics model

The acoustic impedance Z is calculated using Wyllie's equation (Wyllie et al., 1956):

$$Z = V\rho \quad (5.3)$$

where

$$V = \frac{V_r V_f}{\phi V_r + (1 - \phi) V_f} \quad (5.4)$$

and

$$\rho = \phi \rho_f + (1 - \phi) \rho_r \quad (5.5)$$

Here V and ρ denote bulk acoustic velocity and density, V_r , V_f are velocities of the rock matrix and fluid, respectively, and ρ_r and ρ_f are their densities.

The values of V_r and ρ_r depend on the facies type, and typically take constant values, for instance:

$$V_r, \rho_r = \begin{cases} V_1, \rho_1 & \text{if shale} \\ V_2, \rho_2 & \text{if sand} \end{cases} \quad (5.6)$$

We assume that V_r and ρ_r are dependent on both the facies type and porosity in a way that provides the necessary differentiable link for the smooth formulation. We assume that the facies 1 and 2 are characterized by constant porosities ϕ_1 and ϕ_2 , therefore we can suggest the following cubic relationship as a simplified model for the dependence of V_r on ϕ :

$$V_r = a\phi^3 + b\phi^2 + c\phi + d \quad (5.7)$$

where, for instance:

$$\begin{aligned} a &= 4 \frac{V_2 - V_1}{(\phi_2 - \phi_1)^3}, & b &= -\frac{3}{2}a(\phi_1 + \phi_2), \\ c &= -3a \left(\frac{\phi_1 + \phi_2}{2} \right)^2 - b(\phi_1 + \phi_2), & d &= V_1 - a\phi_1^3 - b\phi_1^2 - c\phi_1 \end{aligned} \quad (5.8)$$

The corresponding plot is shown in Figure 5.1, where $V_1 = 3200$ m/s, $V_2 = 2800$ m/s, $\phi_1 = 0.08$, $\phi_2 = 0.26$. Notice, if $\phi = \phi_1$ then $V = V_1$ and, in opposite, if $\phi = \phi_2$ then $V = V_2$.

Substituting in Eq 5.8 V_1 and V_2 with ρ_1 and ρ_2 respectively, we obtain coefficients a' , b' , c' and d' for the model of rock matrix density, which is computed as:

$$\rho_r = a'\phi^3 + b'\phi^2 + c'\phi + d' \quad (5.9)$$

5.2.1.2 Reflectivity series

Reflectivities in depth \mathbf{R}^d are calculated as:

$$R_{i,j}^d = \frac{Z_{i,j+1} - Z_{i,j}}{Z_{i,j+1} + Z_{i,j}}, \quad i = 1, \dots, n_x n_y, \quad j = 1, \dots, n_z \quad (5.10)$$

where surface reflection coefficients are set to zero.

The next step is to map reflectivities represented in depth domain with n_z values into time domain discretized with n_s sampling points as $[t_1 \dots t_{n_s}]$.

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

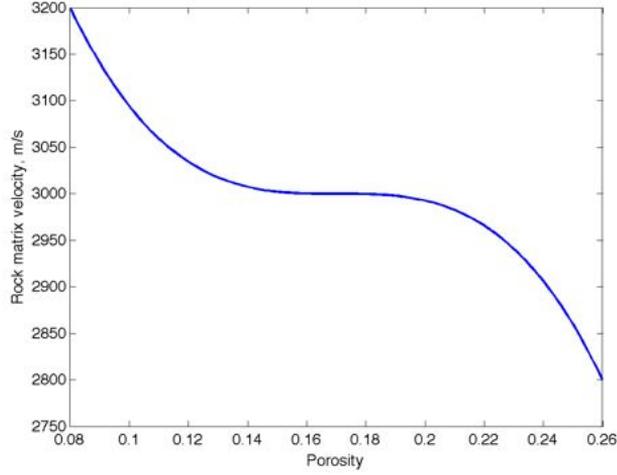


Figure 5.1: Rock matrix velocity V_r as a function of porosity.

Mapping to the sampling point t_s is traditionally performed using two-way travel times via the following procedure:

$$R_{i,s} = R_{i,j^*}^d \quad (5.11)$$

where $j^* : tt_{i,j^*} = \operatorname{argmin}_t t_{i,j}(|t_{i,s} - tt_{i,j}|)$, $j = 1 \cdots n_z$ and the two-way travel time $tt_{i,j} = 2\Delta z \left(\frac{1}{V_{i,1}} + \cdots + \frac{1}{V_{i,j}} \right)$. Unfilled $R_{i,s}$ are taken to be zero.

Notice that this conversion prevents us from differentiation with respect to porosity. To address this difficulty, we propose to define the reflectivity coefficients in time as a linear combination of all reflective coefficients in depth such that,

$$R_{i,s} = e^{-(tt_{i,1}-t_{i,s})^2/\sigma^2} R_{i,1}^d + e^{-(tt_{i,2}-t_{i,s})^2/\sigma^2} R_{i,2}^d + \dots + e^{-(tt_{i,n_z}-t_{i,s})^2/\sigma^2} R_{i,n_z}^d \quad (5.12)$$

The suggested exponential relationship (Figure 5.2) has the following property: when the value of travel time $tt_{i,j}$ is far from the current sampling point $t_{i,s}$, the corresponding reflection coefficient $R_{i,j}$ contributes into Eq. 5.12 with zero, but when $tt_{i,j}$ is close to $t_{i,s}$, its contribution is almost one. Parameter σ is chosen naturally as the sampling interval divided by two.

We compare the proposed approximation of reflectivities in time (5.12) with traditional approach (5.11) in the next section. It is important that now reflection coefficients in time become differentiable with respect to porosity defined in depth since $R = R(tt, R^d)$, $R^d = R^d(Z(\phi))$ and $tt = tt(V(\phi))$.

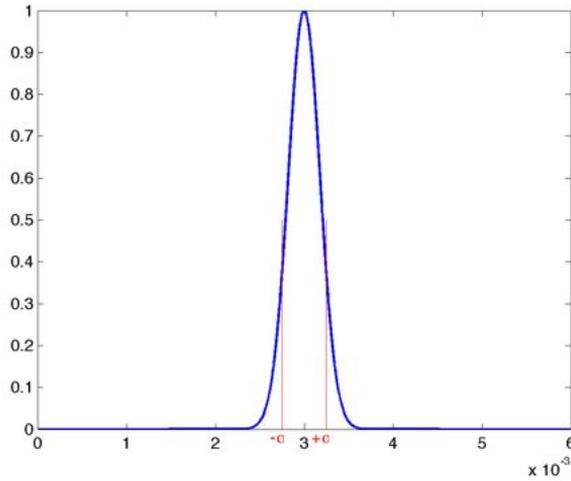


Figure 5.2: The proposed exponential relationship with $t_{i,s} = 0.003$

5.2.2 Seismic sensitivities

The vector of the forward response is shaped by concatenating rows of the matrix \mathbf{S} and therefore consists of $n^{\text{seis}} = n_s n_x n_y$ elements. The Jacobian of the forward response with respect to porosity, per definition, is computed as:

$$\mathbf{G}^{\text{seis}} = \begin{pmatrix} \frac{\partial S_{1,1}}{\partial \phi_{1,1}} & \frac{\partial S_{1,1}}{\partial \phi_{1,2}} & \cdots & \frac{\partial S_{1,1}}{\partial \phi_{n_x n_y, n_z}} \\ \frac{\partial S_{1,2}}{\partial \phi_{1,1}} & \frac{\partial S_{1,2}}{\partial \phi_{1,2}} & \cdots & \frac{\partial S_{1,2}}{\partial \phi_{n_x n_y, n_z}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial S_{n_x n_y, n_z}}{\partial \phi_{1,1}} & \frac{\partial S_{n_x n_y, n_z}}{\partial \phi_{1,2}} & \cdots & \frac{\partial S_{n_x n_y, n_z}}{\partial \phi_{n_x n_y, n_z}} \end{pmatrix} \quad (5.13)$$

Due to the 1D convolution approach, the seismogram measured at the surface location i depends only on porosity values below. This means that the Jacobian matrix is extremely sparse. Fortunately, in quasi-Newton methods, the Jacobian is only used when its product with a data misfit vector is computed (see Sec. 1.5.4), therefore storage of the full Jacobian is not needed. Below we derive the corresponding partial derivatives.

From Eq.5.2 it reads:

$$\left[\frac{\partial S_{i,1}}{\partial \phi_{i,j}} \cdots \frac{\partial S_{i,n_s}}{\partial \phi_{i,j}} \right]^T = W \left[\frac{\partial R_{i,1}}{\partial \phi_{i,j}} \cdots \frac{\partial R_{i,n_s}}{\partial \phi_{i,j}} \right]^T \quad (5.14)$$

where $i = 1, \dots, n_x n_y$ and $j = 1, \dots, n_z$.

From Eq. 5.12 we derive:

$$\frac{\partial R_{i,s}}{\partial \phi_{i,j}} = \sum_{k=1}^{n_z} \frac{\partial L_{i,s,k}}{\partial \phi_{i,j}} R_{i,k}^d + \sum_{k=1}^{n_z} \frac{\partial R_{i,k}^d}{\partial \phi_{i,j}} L_{i,s,k} \quad (5.15)$$

where $L_{i,s,k} = e^{-(t_{i,k} - t_{i,s})^2 / \sigma^2}$

From the definition of two-way travel time we obtain:

$$\frac{\partial L_{i,s,k}}{\partial \phi_{i,j}} = \begin{cases} \frac{4\Delta z(tt_{i,k}-t_{i,s})}{\sigma^2} \frac{\partial V_{i,j}}{\partial \phi_{i,j}} \frac{L_{i,s,k}}{V_{i,j}^2} & \text{if } j < k \\ 0 & \text{if } j > k \end{cases} \quad (5.16)$$

Computation of $\partial V_{i,j}/\partial \phi_{i,j}$ and $\partial R_{i,k}^d/\partial \phi_{i,j}$ is performed by applying the chain rule to Eqs. 5.3 – 5.10 and not shown here.

5.3 Numerical example

We consider a synthetic reservoir model similar to one presented in (Zunino et al. (2013), Appendix I), which was inspired by the Stanford VI reservoir model described in (Lee and Mukerji, 2012). We upscaled the first 80 layers of the original facies model and assumed presence of only two facies. As it was mentioned above, we assume a constant porosity within each facies. We choose shale facies to be defined by a porosity value of 0.08 and sand facies by 0.26. The corresponding true porosity model is shown in Figure 5.3. It consists of 20 horizontal layers, each of size 38 by 38 pixels.

The parameters of the rock physics model are given in Table 5.1. The values of physical properties were inspired by the Stanford VI dataset.

Table 5.1: Parameters of rock physics model

Shale	Sand	Fluid
$V_1 = 3200 \text{ m/s}$	$V_2 = 2800 \text{ m/s}$	$V_f = 1500 \text{ m/s}$
$\phi_1 = 0.08$	$\phi_2 = 0.26$	–
$\rho_1 = 2500 \text{ kg/m}^3$	$\rho_2 = 2650 \text{ kg/m}^3$	$\rho_f = 900 \text{ kg/m}^3$

Table 5.2 lists parameters for modeling the seismic response.

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

Table 5.2: Parameters of the seismic model

Number of sampling points, n_s	120
Sampling interval, Δt	0.0005 s
Wavelet maximum frequency	60 Hz
Total thickness of the formation	60 m

First, in order to test the implementation, we model both the observed data and the forward response without conversion from depth to time. Then we invert noise-free data to obtain porosity. We choose the starting guess randomly in the range between ϕ_1 and ϕ_2 (Fig. 5.4). The result of inversion is shown in Fig. 5.5. As expected, the channels are well reproduced.

Next, we model seismic observations by applying the traditional procedure of conversion from depth to time (Eq. 5.11) and adding Gaussian noise. However, when modeling the forward response, we stay in the depth domain. This results in an incorrect reproduction of geological features (Fig. 5.6).

Figure 5.7 compares seismograms computed using the traditional depth-to-time conversion and the approach described by Eq. 5.12. The similarity is very convincing, therefore we proceed by inverting seismic data (that were computed using Eq.5.11), but modeling the forward response through the procedure described in Eq. 5.12 . The result is shown in Fig. 5.8. Reconstructed porosity is very smooth, however, most of the channels are located in correct locations. Therefore, this solution becomes a very good starting guess for inverting seismic data in combination with multiple point geostatistics. We have used every fifth layer of the true model to construct the training image. The result of joint inversion is shown in Fig 5.9. We have used a one-dimensional template of size 36 x 1 pixels, which showed the best result. The obtained solution came as close as possible to matching the seismic data and reproducing the multiple-point statistics of the training

image.

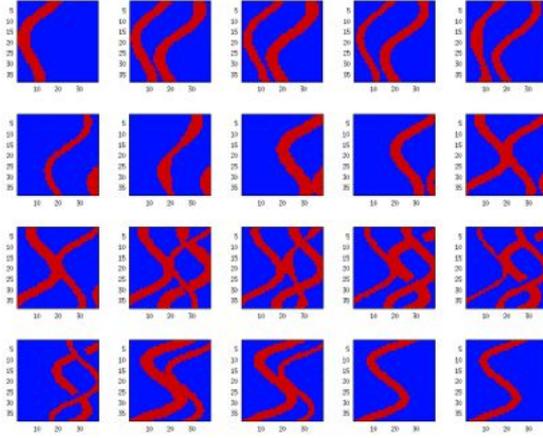


Figure 5.3: True porosity model.

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

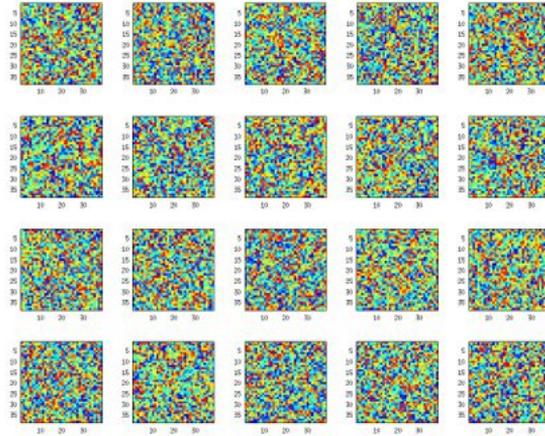


Figure 5.4: Initial randomly generated porosity model.

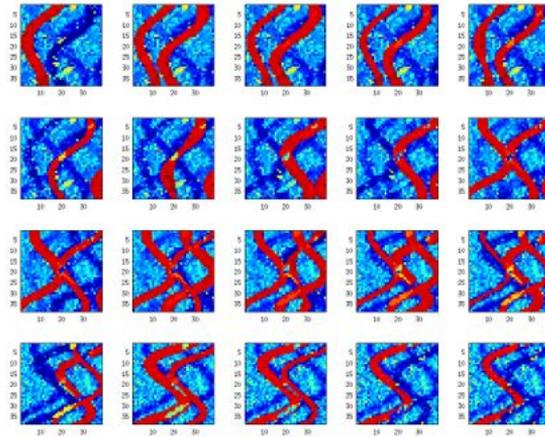


Figure 5.5: Result of inversion when observations and forward model are computed in the depth domain.

5.3. Numerical example

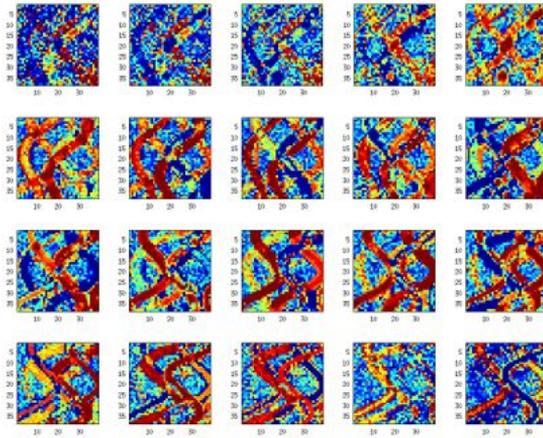


Figure 5.6: Result of inversion, when only forward response is modeled in depth domain.

5. INVERSION OF SEISMIC REFLECTION DATA USING THE SMOOTH FORMULATION

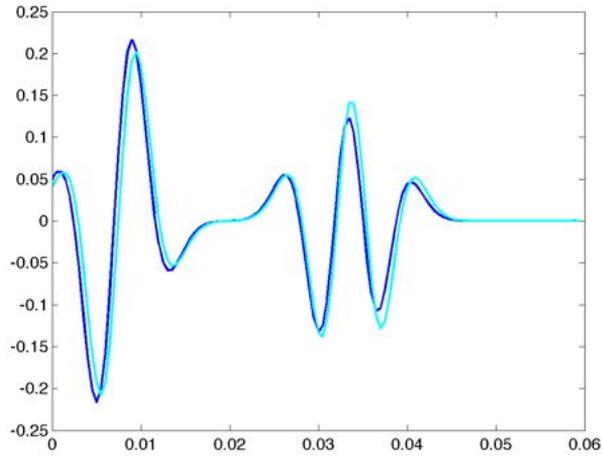


Figure 5.7: Seismograms computed by use of traditional depth-to-time conversion (dark blue) and the proposed procedure (light blue).

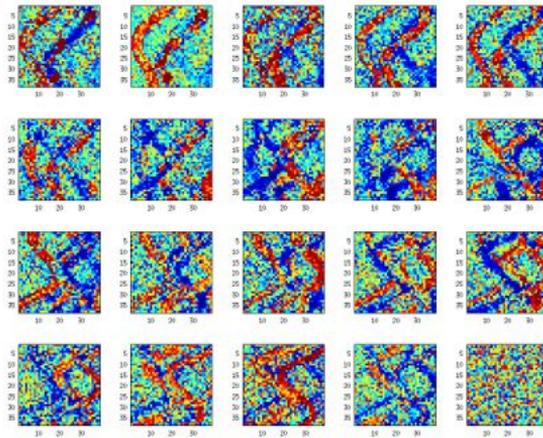


Figure 5.8: Result of inversion with the proposed mapping technique.

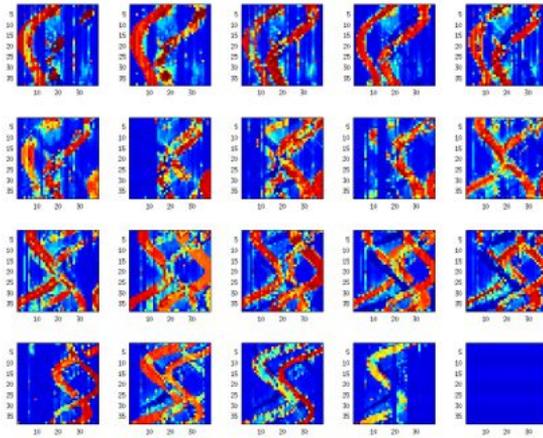


Figure 5.9: Inversion of seismic data with the proposed mapping technique and multiple-point geostatistics.

CHAPTER 6

Conclusion

In this thesis we have discussed challenges of the inverse problems arising in geoscience. Most of them can be described as non-linear, underdetermined and computationally expensive. While inverse problem theory provides us with all necessary tools for consistent integration of information, yet there are practical issues associated with their application.

In this work we aimed at finding a balance between consistent data integration and computational costs. As the result, we proposed an efficient, probabilistically formulated, optimization scheme for solving inverse problems. Integration of complex a priori information, dictated by the inverse problem theory, has a useful property of shrinking solution space and decreasing computational load. The proposed smooth formulation of the multiple-point geostatistics enabled us to use efficient gradient-based optimization. We demonstrated its applicability on the problems of history matching and seismic inversion. The method has a potential to be used in the tasks of closed-loop reservoir management.

The developed method provides us with possibility of detecting different

6. CONCLUSION

islands of high-posterior probabilities, that can be explored by sampling techniques to quantify uncertainty. However, it will be just a small part of the global task of uncertainty quantification, associated with the noise in observations and modelization error. In reservoir characterization uncertainties arise at every stage: at the stage of data acquisition, their processing and interpretation, at the stage of choosing parameterization technique, at the stage of deciding which forward model is to be used. Estimation and propagation of such uncertainties is without doubt one of the most interesting topics in the field of inverse problems.

Bibliography

- Aanonsen, S., Naevdal, G., Oliver, D. S., Reynolds, A. C., and Valles, B. (2009). Ensemble kalman filter in reservoir engineering: a review. *SPE Journal*, 14(3):393–412.
- Arpat, G. B. and Caers, J. (2007). Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203.
- Arroyo-Negrete, E., Devegowda, D., Datta-Gupta, A., and Choe, J. (2008). Streamline-assisted ensemble kalman filter for rapid and continuous reservoir model updating. *SPE Reservoir Evaluation & Engineering*, 11(6):1046–1060.
- Avseth, P., Dvorkin, J., Mavko, G., and Rykkje, J. (2000). Rock physics diagnostic of north sea sands: Link between microstructure and seismic properties. *Geophysical Research Letters*, 27:2761–2764.
- Aziz, K., Durlofsky, L. J., and Tchelepi, H. (2005). *Notes on Petroleum Reservoir Simulation*.
- Bianco, A., Cominelli, A., Dovera, L., Naevdal, G., and Valles, B. (2007). History matching and production forecast uncertainty by means of the ensemble kalman filter: a real field application. In *SPE Europec/EAGE Annual Conference and Exhibition*.
- Bosch, M., Mukerji, T., and Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, 75:A165–A176.

BIBLIOGRAPHY

- Chen, S. F. and Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Comput Speech Lang*, 13:359–394.
- Chen, Y. and Oliver, D. S. (2010a). Cross-covariances and localization for enkf in multiphase flow data assimilation. *Computational Geosciences*, 14:579–601.
- Chen, Y. and Oliver, D. S. (2010b). Ensemble-based closed-loop optimization applied to brugge field. *SPE Reservoir Evaluation & Engineering*, 1:56–71.
- Cordua, K. S., Hansen, T. M., Lange, K., Frydendall, J., and Mosegaard, K. (2012a). Improving multiple-point-based a priori models for inverse problems by combining sequential simulation with the frequency matching method. In *82th Ann. Meet. Soc. Expl. Geophys.*
- Cordua, K. S., Hansen, T. M., and Mosegaard, K. (2012b). Monte carlo full waveform inversion of crosshole gpr data using multiple-point geostatistical a priori information. *Geophysics*, 77:H19–H31.
- Datta-Gupta, A., Kulkarni, K., and Yoon, S. and Vasco, D. (2001). Streamlines, ray tracing and production tomography: generalization to compressible flow. *Petroleum Geoscience*, 7:S75–S86.
- De Vries, L., Carrera, J., Falivene, O., Gratacos, O., and Sloaten, L. (2009). Application of multiple point geostatistics to non-stationary images. *Mathematical Geosciences*, 41:29–42.
- Demyanov, V., Christie, M., Kanevski, M., and Pozdnoukhov, A. (2012). Reservoir modelling under uncertainty - a kernel learning approach. In *Ninth International Geostatistics Congress, Oslo, Norway June 11 Ð 15, 2012*.
- Demyanov, V., Pozdnoukhov, A., Kanevski, M., and Chri (2008). Geomodelling of a fluvial system with semi-supervised support vector regression. In *GEOSTATS 2008, Santiago, Chile*.

- Emerick, A. and Reynolds, A. C. (2011). Combining sensitivities and prior information for covariance localization in the ensemble kalman filter for petroleum reservoir applications. *Computational Geosciences*, 15:251–269.
- Emerick, A. and Reynolds, A. C. (2013). History-matching production and seismic data in a real field using the ensemble smoother with multiple data assimilation. In *2013 SPE Reservoir Simulation Symposium, Feb 18 - 20, 2013 2013, The Woodlands, TX, USA*.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162.
- 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. M. (1993). Multivariate geostatistics: beyond bivariate moments. In *Soares, A. (ed.) Geostatistics Troia*, volume 1. Kluwer Academic.
- Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential gibbs sampling. *Comput Geosci*, 16:593–611.
- Hansen, T. M., Mosegaard, K., and Cordua, K. S. (2008). Using geostatistics to describe complex a priori information for inverse problems. In *Geostatistics 2008*.
- Haugen, V., Naevdal, G., Natvik, L., Evensen, G., Berg, A. M., and Flornes, K. (2008). History matching using the ensemble kalman filter on a north sea field case. *SPE Journal*, 13:382–391.
- Honarkhah, M. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42:487–517.

BIBLIOGRAPHY

- Honarkhah, M. (2011). *Stochastic simulation of patterns using distance-based pattern modeling*. PhD thesis, Stanford University.
- Huang, T., Lu, D. T., Li, X., and Wang, L. (2013). Gpu-based snesim implementation for multiple-point statistical simulation. *Computers & Geosciences*, 54:75–87.
- Jafarpour, B. and Khodabakhshi, M. (2011). A probability conditioning method (pcm) for nonlinear flow data integration into multipoint statistical facies simulation. *Math Geosci*, 43:133–164.
- Lange, K., Frydendall, J., Cordua, K. S., Hansen, T. M., Melnikova, Y., and Mosegaard, K. (2012a). A frequency matching method: solving inverse problems by use of geologically realistic prior information. *Mathematical Geosciences*, 44:783–803.
- Lange, K., Frydendall, J., Cordua, K. S., Hansen, T. M., Melnikova, Y., and Mosegaard, K. (2012b). A frequency matching method: solving inverse problems by use of geologically realistic prior information. *Math Geosci*, 44:783–803.
- Lee, J. and Mukerji, T. (2012). The stanford vi-e reservoir: A synthetic data set for joint seismic-em time-lapse monitoring algorithms. Technical report, Stanford Center for Reservoir Forecasting, Stanford University.
- Lorentzen, R., Flornes, K., and Naevdal, G. (2012). History matching channelized reservoirs using the ensemble kalman filter. *SPE Journal*, 17:137–151.
- Lorentzen, R., Naevdal, G., and Shafieirad, A. (2013). Estimating facies fields by use of the ensemble kalman filter and distance functions—applied to shallow-marine environments. *SPE Journal*, 18:146–158.
- Mariethoz, G., Renard, P., and Caers, J. (2010a). Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research*, 46:W11530.

- Mariethoz, G., Renard, P., and Straubhaar, J. (2010b). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46:W11536.
- Marler, R. T. and Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim*, 26:369–395.
- Melnikova, Y., Lange, K., Frydendall, J., and Mosegaard, K. (2012). A novel approach for combining multiple-point statistics and production data in reservoir characterization. In *Proceedings of the 74th EAGE Conference & Exhibition, Copenhagen, 2012*.
- Melnikova, Y., Zunino, A., Lange, K., Cordua, K. S., and Mos (2013). Discrete-facies history matching in a smooth formulation. *Mathematical Geosciences*, submitted.
- Metropolis, N., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 1(6):1087–1092.
- Mosegaard, K. (1998). Resolution analysis of general inverse problems through inverse monte carlo sampling. *Inverse Problems*, 14:405–426.
- Mosegaard, K. and Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *J. Geophys. Res.*, 100:12431–12447.
- Nocedal, J. and Wright, S. J. (2006). *Numerical optimization*. Springer.
- Oliver, D. S. and Chen, Y. (2011). Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15:185–221.
- Oliver, D. S., Reynolds, A. C., Bi, Z., and Abacioglu, Y. (2001). Integration of production data into reservoir models. *Petroleum Geoscience*, 7:S65–S73.
- Oliver, D. S., Reynolds, A. C., and Liu, N. (2008). *Petroleum reservoir characterization and history matching*. Cambridge University Press, New York.

BIBLIOGRAPHY

- Peaceman, D. (1977). Interpretation of well-block pressures in numerical reservoir simulation. In *SPE-AIME 52 nd Annual Fall Technical Conference and Exhibition, Denver, Oct 9-12, 1977*.
- Sarma, P. and Chen, W. (2009). Generalization of the ensemble kalman filter using kernels for nongaussian random fields. In *SPE Reservoir Simulation Symposium, 2-4 February 2009, The Woodlands, Texas*.
- Sarma, P., Durlofsky, L. J., and Aziz, K. (2008). Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Math Geosci*, 40:3–32.
- Schlumberger GeoQuest, . . (2009). *ECLIPSE reservoir simulator. Technical description*.
- Sech, R., Jackson, M. D., and Hampson, G. J. (2009). Three-dimensional modeling of a shoreface-shelf parasequence reservoir analog: Part 1. surface-based modeling to capture high-resolution facies architecture. *AAPG Bulletin*, 93:1155–1181.
- Spall, J. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Contr*, 37:332–341.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol*, 34(1):1–20.
- Tan, X., Tahmasebi, P., and Caers, J. (2013). Comparing training-image based algorithms using an analysis of distance. *Mathematical Geosciences*, DOI 10.1007/s11004-013-9482-1.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia.

- Van Doren, J. F., Van den Hof, P. M., Jansen, J. D., and Bosgra, O. H. (11421-11426). Determining identifiable parameterizations for large-scale physical models in reservoir engineering. In *Proceedings of the 17th IFAC World Congress, 2008*.
- Vasco, D., Yoon, S., and Datta-Gupta, A. (1999). Integrating dynamic data into high-resolution reservoir models using streamline-based analytic sensitivity coefficients. *SPE J*, 4:389–399.
- Wyllie, M., Gregory, A. R., and Gardner, L. (1956). Elastic wave velocities in heterogeneous and porous media. *Geophysics*, 21:41–70.
- Zhang, F., Reynolds, A. C., and Oliver, D. S. (2003). An initial guess for the levenberg-marquadt algorithm for conditioning a stochastic channel to pressure data. *Mathematical Geology*, 35:67–88.
- Zhang, T., Switzer, P., and Journel, A. G. (2006). Filter based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38:63–80.
- Zunino, A., Lange, K., Melnikova, Y., Hansen, T. M., and Mosegaard, K. (2013). Reservoir modeling combining geostatistics with markov chain monte carlo inversion. In *Mathematics of Planet Earth*, pages 703–707. Springer Berlin Heidelberg.

Appendices

Methods of unconstrained optimization

Algorithms for unconstrained optimization include steepest-descent, Newton, Gauss-Newton, Levenberg-Marquardt and quasi-Newton methods. All the methods require a starting point x_0 that is iteratively updated such that the value of the objective function is minimized. All these methods vary on the strategy for finding the search direction p_k and step length α :

$$x_{k+1} = x_k + \alpha_k p_k \tag{A.1}$$

Steepest descent method, for instance, takes the path along $p_k = -\nabla f_k$, however it does not take into account second derivatives of the function (Hessian) and therefore require a large number of iterations until the convergence.

Newton method, in opposite has the highest convergence speed, however, the knowledge of the Hessian is needed:

$$p_k = -(\nabla^2 f_k)^{-1} \nabla f_k \tag{A.2}$$

It can be computationally unfeasible to compute Hessian for large scale nonlinear problems. Therefore other methods use approximations of the second derivatives. Gauss-Newton method that was developed specifically for solving least-squares problems:

$$f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x) \quad (\text{A.3})$$

$$J(x) = \left[\frac{\partial r_j}{\partial x_i} \right]_{j=1, \dots, m \quad i=1, \dots, n} \quad (\text{A.4})$$

Then

$$\nabla f(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x) \quad (\text{A.5})$$

And consequently, the Hessian:

$$\begin{aligned} \nabla^2 f(x) &= \sum_{j=1}^m \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) \\ &= J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) \end{aligned} \quad (\text{A.6})$$

In Gauss-Newton and Levenberg-Marquardt techniques we neglect with the last term in Eq.A.6 and therefore the Hessian is completely defined by the first derivatives.:

$$\nabla f^2(x) = J(x)^T J(x) \quad (\text{A.7})$$

Therefore, essentially, the Gauss-Newton method is the Newton method with the approximated Hessian. Levenberg-Marquardt in its own turn, can be viewed as the Gauss-Newton method with the trust region strategy for finding the search direction. Levenberg-Marquardt typically requires more iterations to converge, however it is more robust.

These methods are very popular, since they do not require the knowledge of Hessian and much more faster than usual steepest-descent. The complication arise usually when the size of parameters in the model is very high.

Quasi-Newton methods also use approximation of the Hessian. In general, the search direction is defined as follows:

$$p_k = -B_k^{-1} \nabla f_k \tag{A.8}$$

where B_k is the approximation of the Hessian at the k 'th iteration. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula dictates the iterative update:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{A.9}$$

where $s_k = x_{k+1} - x_k = \alpha_k p_k$ and $y_k = \nabla f_{k+1} - \nabla f_k$.

In addition these methods do not require explicit knowledge of the Jacobian.

Defining prior probabilities: connection with chi-square and multivariate Gaussian

Consider two discrete images: test and training. Let us assume that their frequency distributions consist of 3 categories only. The number of counts is equal to N . Then the histogram of the test image is $\{H_1, H_2, H_3\}$ and the histogram of the training image $\{Np_1, Np_2, Np_3\}$. Assuming that the frequency distribution of the training image is the underlying theoretical distribution, we define the chi-square distance:

$$\chi^2 = \frac{(H_1 - Np_1)^2}{Np_1} + \frac{(H_2 - Np_2)^2}{Np_2} + \frac{(H_3 - Np_3)^2}{Np_3} \quad (\text{B.1})$$

Since $H_1 + H_2 + H_3 = N$ and $p_1 + p_2 + p_3 = 1$, we have:

$$\chi^2 = \frac{(H_1 - Np_1)^2}{Np_1} + \frac{(H_2 - Np_2)^2}{Np_2} + \frac{(N - H_1 - H_2 - N(1 - p_1 - p_2))^2}{N(1 - p_1 - p_2)} \quad (\text{B.2})$$

B. DEFINING PRIOR PROBABILITIES: CONNECTION WITH CHI-SQUARE
AND MULTIVARIATE GAUSSIAN

Then

$$\chi^2 = \frac{(H_1 - Np_1)^2}{Np_1} + \frac{(H_2 - Np_2)^2}{Np_2} + \frac{(H_1 + H_2 - Np_1 - Np_2)^2}{N(1 - p_1 - p_2)} \quad (\text{B.3})$$

After regrouping the equation looks as follows:

$$\chi^2 = \frac{1}{Np_1p_2(1 - p_1 - p_2)} \left((H_1 - Np_1)^2 p_2(1 - p_2) + \right. \\ \left. (H_2 - Np_2)^2 p_1(1 - p_1) + 2(H_1 - Np_1)(H_2 - Np_2)p_1p_2 \right) \quad (\text{B.4})$$

Denoting $z = \frac{1}{Np_1p_2(1 - p_1 - p_2)}$ it is easy to write the equation in matrix form:

$$\chi^2 = z \begin{bmatrix} H_1 - Np_1 & H_2 - Np_2 \end{bmatrix} \begin{bmatrix} p_2(1 - p_2) & p_1p_2 \\ p_1p_2 & p_1(1 - p_1) \end{bmatrix} \begin{bmatrix} H_1 - Np_1 \\ H_2 - Np_2 \end{bmatrix} \quad (\text{B.5})$$

Let us define as \mathbf{M} the following matrix:

$$\mathbf{M} = z \begin{bmatrix} p_2(1 - p_2) & p_1p_2 \\ p_1p_2 & p_1(1 - p_1) \end{bmatrix} \quad (\text{B.6})$$

We assume that $\mathbf{M} = \mathbf{C}^{-1}$, where \mathbf{C} is the covariance matrix. Then $\mathbf{C} = \mathbf{M}^{-1}$:

$$\mathbf{C} = \frac{Np_1p_2(1 - p_1 - p_2)}{p_2(1 - p_2)p_1(1 - p_1) - p_1^2p_2^2} \begin{bmatrix} p_1(1 - p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1 - p_2) \end{bmatrix} \quad (\text{B.7})$$

And even shorter:

$$\mathbf{C} = \begin{bmatrix} Np_1(1 - p_1) & -Np_1p_2 \\ -Np_1p_2 & Np_2(1 - p_2) \end{bmatrix} \quad (\text{B.8})$$

Now we are ready to compute the prior probability:

$$f(H_1, H_2, H_3) = f(H_1, H_2) = \text{const} \cdot \exp\left(-\frac{1}{2}\chi^2\right) \quad (\text{B.9})$$

Notice the decrease in the degrees of freedom.

The PDF for the *non-degenerate* multivariate normal distribution has the following form:

$$f_{\mathbf{x}}(x_1, \dots, x_s) = \frac{1}{(2\pi)^{s/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (\text{B.10})$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$ and $s = \text{rank}(\boldsymbol{\Sigma})$.

Since our 3-variate case was degraded to bivariate, we have $\text{rank}(\mathbf{C}) = 2$:

$$\begin{aligned} f_{\mathbf{H}}(H_1, H_2, H_3) &= f_{\mathbf{H}}(H_1, H_2) \\ &= \frac{1}{(2\pi)^{2/2} |\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{H}^* - N\mathbf{p}^*)^T \mathbf{C}^{-1}(\mathbf{H}^* - N\mathbf{p}^*)\right) \end{aligned} \quad (\text{B.11})$$

where $\mathbf{H}^* = (H_1, H_2)$ and $\mathbf{p}^* = (p_1, p_2)$.

Equation B.8 defines the covariance matrix \mathbf{C} , therefore:

$$|\mathbf{C}| = N^2 p_1 p_2 (1 - p_1 - p_2) = N^2 p_1 p_2 p_3 \quad (\text{B.12})$$

Generalization

We can conclude that in the general case the following is true:

$$\begin{aligned} f_{\mathbf{H}}(H_1, \dots, H_k) &= f_{\mathbf{H}}(H_1, \dots, H_{k-1}) \\ &= \frac{1}{(2\pi)^{(k-1)/2} |\mathbf{C}^*|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{H}^* - N\mathbf{p}^*)^T \mathbf{C}^{*-1}(\mathbf{H}^* - N\mathbf{p}^*)\right) \end{aligned} \quad (\text{B.13})$$

B. DEFINING PRIOR PROBABILITIES: CONNECTION WITH CHI-SQUARE
AND MULTIVARIATE GAUSSIAN

where $\mathbf{H}^* = (H_1, \dots, H_{k-1})$ and $\mathbf{p}^* = (p_1, \dots, p_{k-1})$.
 \mathbf{C}^* is $k-1 \times k-1$ covariance matrix, where:

$$\begin{aligned} C_{ii}^* &= Np_i(1-p_i), \\ C_{ij}^* &= -Np_i p_j, \quad i = 1, \dots, k-1, \quad j = 1, \dots, k-1 \end{aligned} \quad (\text{B.14})$$

and

$$|\mathbf{C}^*| = N^{k-1} p_1 p_2 \cdots p_k \quad (\text{B.15})$$

Ricker Wavelet

Amplitudes of the Ricker wavelet of maximum frequency w^{\max} at time t is computed as :

$$w = (1 - 2(\pi w^{\max} t)^2) e^{-(\pi w^{\max} t)^2} \quad (\text{C.1})$$

We consider the wavelet computed at the following discrete moments of time :

$$[w_1 \cdots w_{n_s}] = [w(t_1) \cdots w(t_{n_s})] \quad (\text{C.2})$$

where $[t_1, \cdots, t_{n_s}] = [0, \Delta t, 2\Delta t, \cdots, (n_s - 1)\Delta t]$ and Δt is the sampling interval.

Since the wavelet is symmetric around zero,

$$[w(t_1) \cdots w(t_{n_s})] = [w(t_{-1}) \cdots w(t_{-n_s})], \text{ where}$$

$$[t_{-1}, \cdots, t_{-n_s}] = [0, -\Delta t, -2\Delta t, \cdots, -(n_s - 1)\Delta t].$$

To compute convolution, the following operator is constructed:

$$\mathbf{W} = \begin{pmatrix} w_1 & w_{-2} & \cdots & w_{-n_s} \\ w_2 & w_1 & \cdots & w_{-n_s+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_s} & w_{n_s-1} & \cdots & w_1 \end{pmatrix} \quad (\text{C.3})$$

And due to the aforementioned symmetry:

$$\mathbf{W} = \begin{pmatrix} w_1 & w_2 & \cdots & w_{n_s} \\ w_2 & w_1 & \cdots & w_{n_s-1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_s} & w_{n_s-1} & \cdots & w_1 \end{pmatrix} \quad (\text{C.4})$$

APPENDIX D I

Paper 1

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

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
© The Author(s) 2012. This article is published with open access at Springerlink.com

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}^{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,

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

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}^{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}^{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

$$\rho(\mathbf{m}) = \text{const.} \exp(-\alpha f(\mathbf{m})),$$

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,

$$L(\mathbf{m}, \mathbf{d}^{\text{obs}}) = \text{const.} \exp\left(-\frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{\mathbf{C}_d}^2\right).$$

Here, \mathbf{C}_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

$$\sigma(\mathbf{m}) = \text{const.} \rho(\mathbf{m}) L(\mathbf{m}, \mathbf{d}^{\text{obs}}).$$

The set of model parameters that maximizes the a posteriori probability density is called the maximum a posteriori (MAP) model

$$\begin{aligned} \mathbf{m}^{\text{MAP}} &= \arg \max_{\mathbf{m}} \{\sigma(\mathbf{m})\} \\ &= \arg \min_{\mathbf{m}} \{-\log \sigma(\mathbf{m})\} \\ &= \arg \min_{\mathbf{m}} \left\{ \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{\mathbf{C}_d}^2 + \alpha f(\mathbf{m}) \right\}. \end{aligned}$$

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 $\|\mathbf{m} - \mathbf{m}^{\text{prior}}\|_{\mathbf{C}_m}^2$, stemming from a Gaussian a priori distribution of the model parameters with mean values $\mathbf{m}^{\text{prior}}$ and covariance matrix \mathbf{C}_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 $Z = \{1, 2, \dots, N\}$ with N voxels (or pixels if the image is only two dimensional) where the voxels can have the m different values $0, 1, \dots, m - 1$. We introduce the N variables, z_1, z_2, \dots, z_N and let z_k describe the value of the k th voxel of the image. It is assumed that the image is a realization of an unknown, random process satisfying:

1. The value of the k th 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

$$f_Z(z_k|z_N, \dots, z_{k+1}, z_{k-1}, \dots, z_1) = f_Z(z_k|\mathbf{z}_k),$$

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 $Z = \{1, \dots, N\}$ 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 \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

$$Z_{\text{in}} = \left\{ k \in Z : |\mathcal{N}_k| = \max_{l \in Z} |\mathcal{N}_l| \right\},$$

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 $i = 1, \dots, m^{n+1}$, count the number of patterns that have the i th pattern value. The frequency distribution is then defined as $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = [\pi_1, \dots, \pi_{m^{n+1}}].$$

Let p_Ω denote the mapping from voxel values of an image Z to its frequency distribution $\boldsymbol{\pi}$, that is, $p_\Omega(z_1, \dots, z_N) = \boldsymbol{\pi}$.

Figure 1 shows an example of an image and the patterns it contains for the template function that defines neighborhoods as follows

$$\mathcal{N}_k = \{l \in Z \setminus \{k\} : |l_x - k_x| \leq 1, |l_y - k_y| \leq 1\}.$$

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

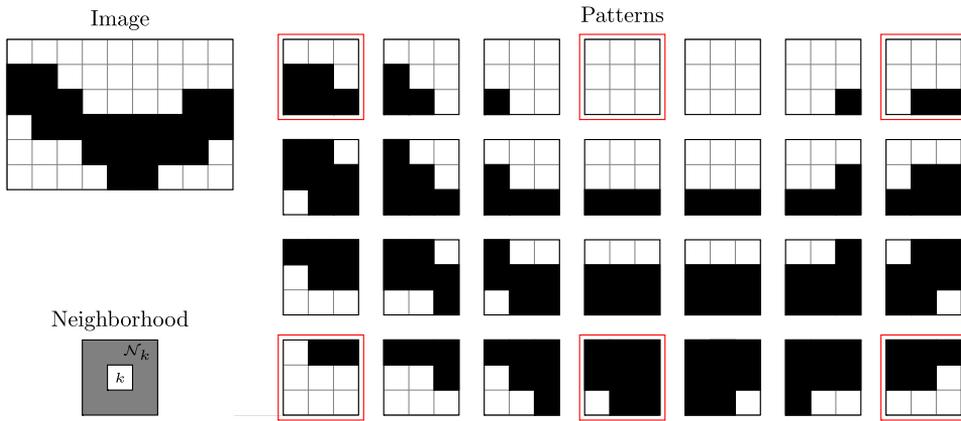


Fig. 1 Example of patterns found in an image. Notice how the image is completely described by the (ordered) patterns in every third row and column; the patterns are marked in red

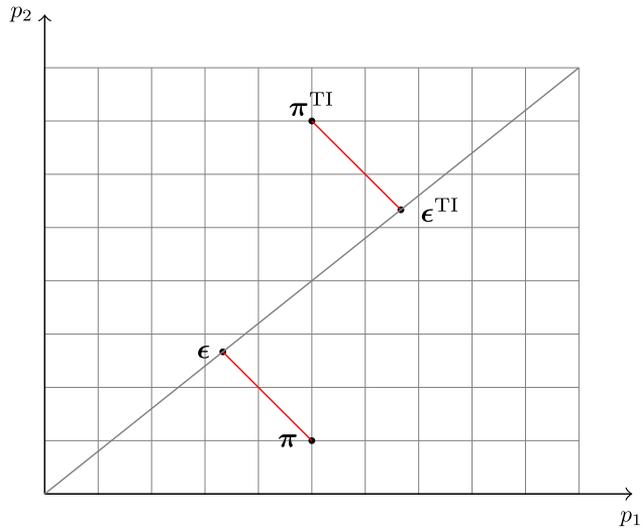
distribution of an image is uniquely determined. The opposite, however, does not hold. Different images can, excluding symmetries, have the same frequency distribution. This is what the FM method seeks to exploit by using the frequency distribution to generate new images, at the same time similar to, 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 π and π^{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 $\|\pi - \epsilon\|_2 = \|\pi^{\text{TI}} - \epsilon^{\text{TI}}\|_2$



Hence, given the frequency distributions of an image, π , and of a training image, π^{TI} , and by letting

$$I = \{i \in \{1, \dots, m^{n+1}\}: \pi_i^{\text{TI}} > 0\} \cup \{i \in \{1, \dots, m^{n+1}\}: \pi_i > 0\}, \tag{2}$$

we compute what we define as the dissimilarity function value of the image

$$c(\pi) = \chi^2(\pi, \pi^{\text{TI}}) = \sum_{i \in I} \frac{(\pi_i^{\text{TI}} - \epsilon_i^{\text{TI}})^2}{\epsilon_i^{\text{TI}}} + \sum_{i \in I} \frac{(\pi_i - \epsilon_i)^2}{\epsilon_i}, \tag{3}$$

where ϵ_i denotes the counts of the underlying distribution of patterns with the i th 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 i th pattern value for images of the same size as the training image. These counts are computed as

$$\epsilon_i = \frac{\pi_i + \pi_i^{\text{TI}}}{n_Z + n_{\text{TI}}} n_Z, \tag{4}$$

$$\epsilon_i^{\text{TI}} = \frac{\pi_i + \pi_i^{\text{TI}}}{n_Z + n_{\text{TI}}} n_{\text{TI}}, \tag{5}$$

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

$$\begin{aligned} & \min_{z_1, \dots, z_N} \left\| \mathbf{d}^{\text{obs}} - g(z_1, \dots, z_N) \right\|_{\mathbf{C}_d}^2 + \alpha c(\boldsymbol{\pi}), \\ & \text{w.r.t. } \boldsymbol{\pi} = p_{\Omega}(z_1, \dots, z_N), \\ & z_k \in \{0, \dots, m-1\} \quad \text{for } k = 1, \dots, N, \end{aligned} \quad (6)$$

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 $z_k \in \{0, \dots, m-1\}$ for $k = 1, \dots, N$ 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}_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 $\{0, \dots, m - 1\}$ 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 π^{TI} and pattern list \mathbf{p} (Algorithm 2)
 Compute partial frequency distribution of starting image π (Algorithm 3)
while *not converged* **do**
 Compute perturbed image \bar{Z} based on Z (Algorithm 4)
 Compute partial frequency distribution of perturbed image $\bar{\pi}$ (Algorithm 5)
 if *accept the perturbed image* **then**
 Set $Z \leftarrow \bar{Z}$ and $\pi \leftarrow \bar{\pi}$
 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 $\boldsymbol{\pi}^{\text{TI}} > 0$. 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

$$\begin{aligned}
 c(\boldsymbol{\pi}) &= \sum_{i \in I} \frac{(\pi_i^{\text{TI}} - \epsilon_i^{\text{TI}})^2}{\epsilon_i^{\text{TI}}} + \sum_{i \in I} \frac{(\pi_i - \epsilon_i)^2}{\epsilon_i} \\
 &= \sum_{i \in I} \frac{(\sqrt{\frac{n_Z}{n_{\text{TI}}}} \pi_i^{\text{TI}} - \sqrt{\frac{n_{\text{TI}}}{n_Z}} \pi_i)^2}{\pi_i^{\text{TI}} + \pi_i}.
 \end{aligned}
 \tag{7}$$

This leads to the introduction of the following two subsets of I

$$\begin{aligned}
 I_1 &= \{i \in I : \pi_i^{\text{TI}} > 0\}, \\
 I_2 &= \{i \in I : \pi_i^{\text{TI}} = 0\}.
 \end{aligned}$$

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

$$\begin{aligned}
 c(\boldsymbol{\pi}) &= \sum_{i \in I_1} \frac{(\sqrt{\frac{n_Z}{n_{\text{TI}}}} \pi_i^{\text{TI}} - \sqrt{\frac{n_{\text{TI}}}{n_Z}} \pi_i)^2}{\pi_i^{\text{TI}} + \pi_i} + \frac{n_{\text{TI}}}{n_Z} \sum_{i \in I_2} \pi_i \\
 &= \sum_{i \in I_1} \frac{(\sqrt{\frac{n_Z}{n_{\text{TI}}}} \pi_i^{\text{TI}} - \sqrt{\frac{n_{\text{TI}}}{n_Z}} \pi_i)^2}{\pi_i^{\text{TI}} + \pi_i} + \frac{n_{\text{TI}}}{n_Z} \left(n_Z - \sum_{i \in I_1} \pi_i \right)
 \end{aligned}
 \tag{8}$$

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 π^{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 π^{TI} .

Algorithm 2: Frequency Distribution of a Training Image

Input: Training Image Z^{TI}
Output: Frequency distribution π^{TI} , list of pattern values \mathbf{p}
 Initialization: empty list π^{TI} , empty list \mathbf{p}
for each inner voxel, i.e., $k \in Z_{\text{in}}^{\text{TI}}$ **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 \mathbf{p}
 | | Set the corresponding new entry of π^{TI} equal to 1
 | **end**
end

Algorithm 3 computes the partial frequency distribution π of an image that is needed to evaluate the dissimilarity function $c(\pi) = \chi^2(\pi, \pi^{\text{TI}})$ 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 \mathbf{p} from the training image
Output: Partial frequency distribution π
 Initialization: all zero list π (same length as \mathbf{p})
for each inner voxel, i.e., $k \in Z_{\text{in}}$ **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 π of Z

Output: Perturbed image \bar{Z}

Initialization: set $\bar{\pi} = \pi$

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 addition 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 \bar{Z} , that is, the set of voxels whose pattern values are changed when perturbing image Z into image \bar{Z}

$$Z^{\text{aff}} = \{k \in Z: p_k \neq \bar{p}_k\}. \quad (9)$$

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 π of Z , perturbed image \bar{Z} , set of affected voxels Z^{aff} , set of pattern values \mathbf{p} from the training image

Output: Partial frequency distribution $\bar{\pi}$ of \bar{Z}

Initialization: set $\bar{\pi} = \pi$

for each affected voxel, i.e., $k \in Z^{\text{aff}}$ **do**

Extract pattern k from both Z and \bar{Z}

Compute both pattern values p_k and \bar{p}_k

if the pattern p_k is present in the training image **then**

| Subtract 1 from the corresponding entry of $\bar{\pi}$

end

if the pattern \bar{p}_k is present in the training image **then**

| Add 1 to the corresponding entry of $\bar{\pi}$

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)

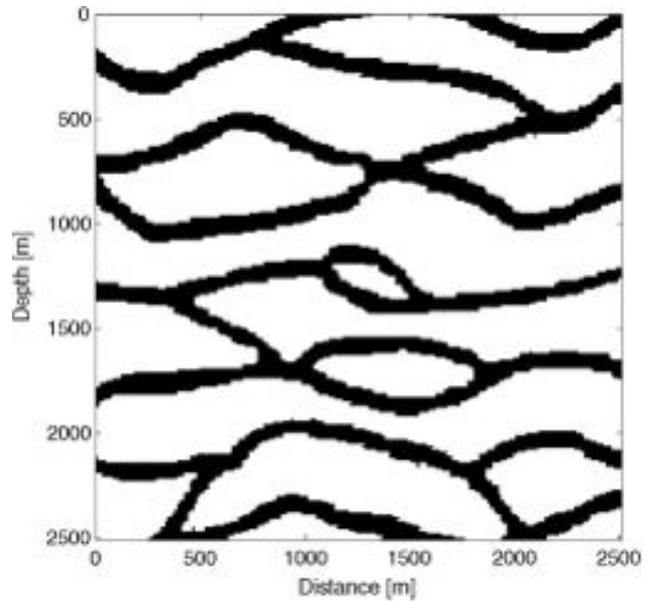


Table 1 Parameter values for
the test case

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)

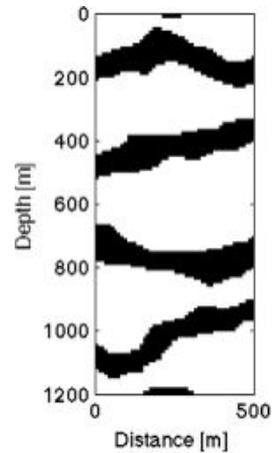
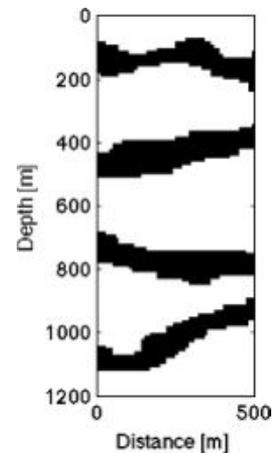


Fig. 5 Computed model for $\alpha = 1.8 \times 10^{-2}$ (resolution: 50×120 pixels)



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

$$\mathcal{N}_k = \{l \in Z \setminus \{k\}: |l_x - k_x| \leq 4, |l_z - k_z| \leq 3\}.$$

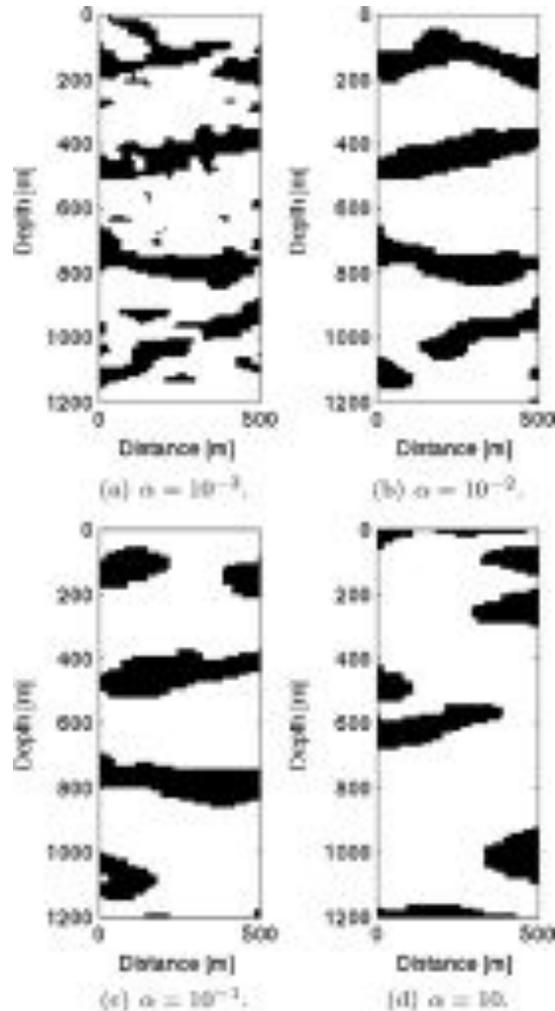
Recall that pixel l has the coordinates (l_x, l_z) ; the first coordinate being the horizontal distance from the left borehole and the second coordinate being the depth, both measured in pixels. To compute a perturbed image, the domain used in Algorithm 4 is defined as follows

$$\mathcal{D}_k = \{l \in Z \setminus Z^{\text{cond}}: |l_x - k_x| \leq 7, |l_z - k_z| \leq 7\}.$$

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

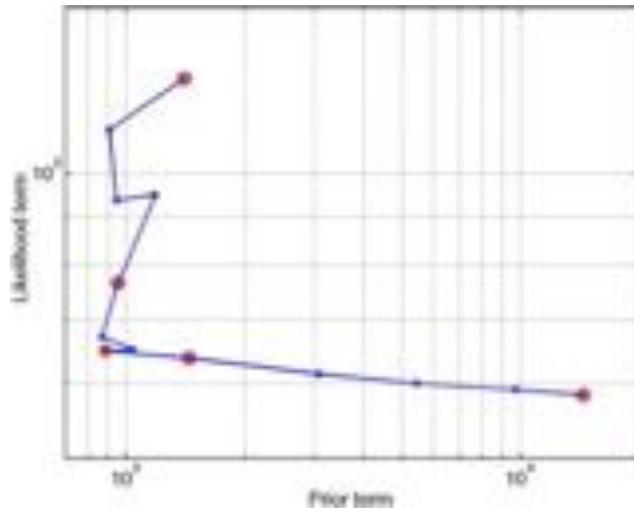


Figure 4 shows the reference model that describes what is considered to be the true velocity profile between the two boreholes. The image has been generated by the *snesim* algorithm (Strebelle 2002) using the multiple point statistics of the training image. The arrival times \mathbf{d} for the reference model \mathbf{m}^{ref} are computed by a forward computation, $\mathbf{d} = G\mathbf{m}^{\text{ref}}$. We define the observed arrival times \mathbf{d}^{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 from 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 $\frac{1}{2} \|g(\mathbf{m}^{\text{FM}}) - \mathbf{d}^{\text{obs}}\|_{C_d}^2$ 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.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

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: *Geostatistics-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 & Hall/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



Paper 2

A Novel Approach for Combining Multiple-Point Statistics and Production Data in Reservoir Characterization

Authors: Yulia Melnikova, Katrine Lange, Jan Frydendall and Klaus Mosegaard

Published in: Proceedings of 74th EAGE Conference & Exhibition Incorporating SPE EUROPEC 2012 Copenhagen, Denmark 4-7 June 2012

Introduction

History matching problem arises regularly at the stage of reservoir development and its performance optimization. One wants to match simulation response with production data adjusting reservoir model parameters, e.g. permeability values, location of faults and fractures. Good history-matched model must possess two important properties: be able to match data observations within their uncertainty and be consistent with geological expectations, i.e. prior information. If most of the attention is paid to the minimization of the data misfit, the geological realism of the solution may suffer. Traditionally, prior information is conserved in the form of covariance model that is defined by two-point statistics. This means that spatial variability is accounted for only on a pairwise basis, and as a result, curvilinear long-correlated features become neglected.

With the development of data assimilation techniques, such as the Ensemble Kalman Filter (EnKF), matching the data and estimation of the uncertainty became a feasible task (Aanonsen et al. 2009). However, due to the mentioned traditional approach of incorporating prior information, geologically realistic features, such as channels are hardly ever reproduced. To keep complex geological structures in the solution, multiple-point statistics framework has to be used (Journal and Zhang 2006). Many authors suggest inferring complex spatial information from the so-called *training images* (Strebelle 2002, Caers 2003, Eskandaridavand 2010). Training images contain expected geological features and can be constructed using geologists' expertise, database of characteristic structures, photographs of the outcrops.

In this study we also use multiple point statistics of training images to provide geological realism of the solution. This paper is the first application of the Frequency Matching (FM) method (Lange et al. 2011) to the solution of the history matching problem. The FM method allows us to accurately quantify the consistency of the model with complex prior, e.g. training image, computing prior probability of the model. Consequently, it enables us to guide the model safely by both prior information and data observations.

Description of the method can be found in the next section. It is followed by a 3D synthetic example and the conclusion.

Methodology

In this study we apply a multiple-point statistics framework defined by the FM method for solving history matching problem. Multiple point statistics gained its popularity in characterization of sedimentary reservoirs that possess channel-like features; see, for example, paper by Hoffman et al. (2006). The Probability perturbation method (PPM) suggested by Caers and Hoffman (2006) aims at finding solution that is consistent both with prior information, i.e. obtained from a training image, and data observations. However, as stated in Hansen et al. (2012), the PPM approach finds the solution that belongs to the space of prior models allowed by training image and only maximizes the data fit. While the FM method is a Bayesian approach and hence maximizes the posteriori model. The FM technique characterizes images by their multiple point statistics. To retrieve multiple point statistics from an image, a scanning template is applied to it. Further, the scanned information is sorted and forms the frequency distribution of the image. In such a way, the image is uniquely described by the histogram of the multi-point spatial event. For comparing of two images, a dissimilarity measure is introduced, and defined as χ^2 distance between their histograms.

Generally, the FM method can be used for the solution of inverse problems, where one wants to estimate a model \mathbf{m} , given some data observations \mathbf{d}_{obs} , with respect to a complex prior information in the form of a training image. Mathematically, this can be formulated as the following optimization problem:

$$\min_{\mathbf{m}} \left\| \mathbf{d}_{\text{obs}} - g(\mathbf{m}) \right\|_{C_d}^2 + \alpha \chi^2(\pi, \pi^{TI}) \quad (1)$$

where g is non-linear forward operator, C_d is data covariance matrix, π and π^{TI} are the frequency distributions of the test image and training image respectively and α is a weight parameter. The first term in equation (1) minimizes the difference between observations and the forward simulation

response, while the second term minimizes the discrepancy between statistics of the model and of the training image.

We solve the optimization problem (1) with a greedy stochastic annealing algorithm. First, we choose a proper scanning template and construct the histograms of the training image and the starting model. Then we conduct a random walk in model space, suggesting change for the values of voxels (pixels in 2D space). The change is accepted or rejected depending on the optimization method criteria. For the test case, described in the next section we used the “greedy” approach: if the suggested change decreased the value of the objective function (1) the change was accepted. The greedy stochastic annealing algorithm may get stuck in local minima. However, it is more computationally efficient and provides sufficiently low values for the both terms in the objective function (1). Since the history matching problem is very much undetermined, we are satisfied with a solution that honours (with desirable accuracy) both data and prior information. The described iterative approach involves one forward simulation per iteration. This is a bottleneck of the method. However, the development of the FM method is an active topic of the research and the strategy for decreasing number of forward simulations is under investigation. For example, increasing amount of flipped blocks may improve the algorithm performance.

Example

In this section we test the Frequency Matching method on the 3D synthetic example. Let us consider a 3D synthetic oil reservoir of 25x25x3 cells. Physical dimension of a cell is 50x50x10 m. Wells configuration is a traditional nine-spot pattern with one water injector in the middle and 8 producers on the sides. We use a streamline simulator for modelling flow response. Initial water saturation is 0, initial pressure 100 bar.

As geological model, we use binary training image of size 60x60x3 with distinct narrow high permeable channels of 500 mD and shale background of 50 mD, as shown in **Figure 1** (all three layers are the same). It should be mentioned, that generally the FM method is suitable for the assessing priors with multiple categorical values. The reference permeability model, shown on **Figure 22** together with the well positions, just as the training image, is presented by two discrete values of 500 mD and 50 mD.



Figure 1 Training image, 60x60x3 cells

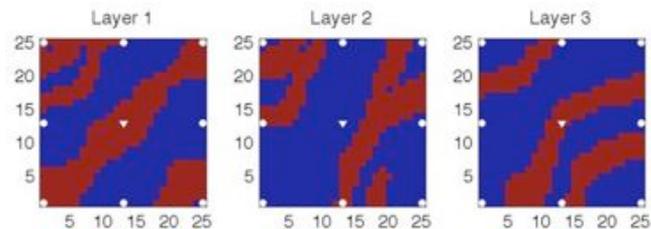


Figure 2 Reference permeability 25x25x3 cells, layers 1-3

Production history was generated applying forward simulation to the reference model and adding five percent Gaussian noise. Observations consist of 5 measurements of oil rate for each producer at 600, 1200, 1800, 2400 and 3000 days respectively. Note that we only gauge the cumulative production value of data at each well, and not at each segment of the wells. The corresponding data covariance matrix is a diagonal matrix with the values equal to the added noise variance.

The starting model consists of random combination of channel and shale facies. We follow the algorithm described in the previous section. It is worth saying that for computational efficiency only a 2D scanning template of 3x3 pixels was used. However, this choice is unlikely to have influence on the result as the reference channels have vertical continuity of one pixel. The value of the weight parameter α from equation (1) was chosen empirically to be equal to 1, however, it is clearly a question for the future research.

The solution model was obtained at the moment of approximately of 30000 iterations. Agreement with the prior is assessed by comparing the histograms of the solution and the training image in Figure 3: at first glance the histograms seems the same, however, there are still some discrepancies. With a larger template or with more iteration we may have improved more on the result. Analysis of the solution (**Figure 4**) shows that some of the features, in comparison to the reference model, were successfully reproduced, for example, the diagonal connection in the first layer. Dissimilarities may be explained by the low amount of the data values, which is unavoidable when dealing with the ill-posed inverse problem.

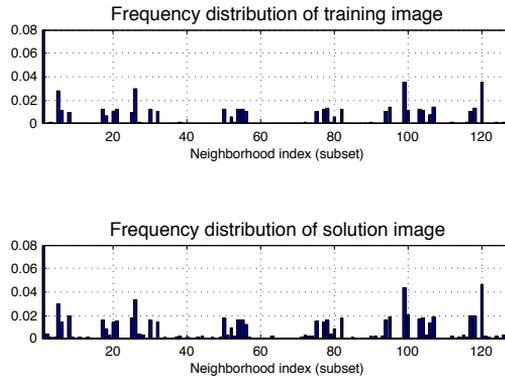


Figure 3 Histograms of the training image (upper) and the model (lower) at the moment of 30000 iterations.

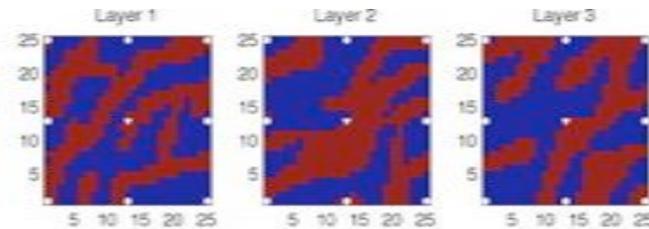


Figure 4 Permeability model at the moment of 30000 iterations

As for the quality of the data match, we can infer from **Figure 5** and Figure 6 that most of the wells were matched within their uncertainty. The legend is following: red solid line - data with noise, red dashed line - data without noise, blue line with diamond marker – data from the solution.

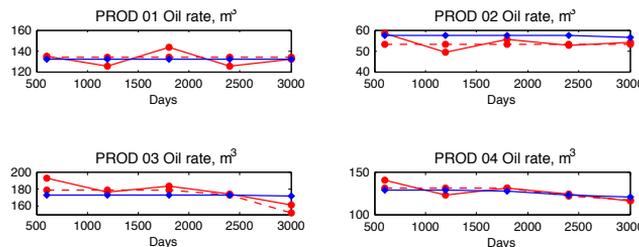


Figure 5 Oil production rate for wells 1-4.

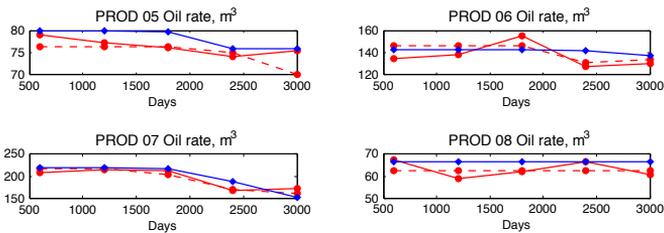


Figure 6 Oil production rate for wells 4-8.

Conclusions

We demonstrated a new multiple-points statistics framework for finding geology-consistent solution of history matching problem. We used the Frequency Matching method that allows us to combine prior information based on training image and production data. The 3D synthetic test case showed that the obtained solution was consistent both with prior information and data observations. This demonstrated the potential of the method and suggests that it could be used on more complicated cases. Future work will be related on improving of computational efficiency

References

- Anonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., and Vallès, B. [2009] The ensemble Kalman filter in reservoir engineering – a review. *SPE Journal*, **14**(3), 393-412.
- Caers, J. [2003] History matching under a training image-based geological model constraint. *SPE Journal*, **8**(3), 218–226.
- Caers, J. and Hoffman T. [2006] The probability perturbation method: A new look at bayesian inverse modeling. *Mathematical Geology*, **38**(1), 81–100.
- Eskandaridavand K. and Srinivasan S. [2010] Reservoir modelling of complex geological systems – a multiple-point perspective, *Journal of Canadian Petroleum Technology*, **49**(8), 59-68.
- Hansen, T. M., Cordua, K. S., and Mosegaard, K., [2012] Inverse problems with non-trivial priors - Efficient solution through Sequential Gibbs Sampling. *Computational Geosciences*, doi:10.1007/s10596-011-9271-1
- Hoffman, T., Caers, J., Wen, X. and Strebelle, S. [2006] A Practical Data-Integration Approach to History Matching: Application to a Deepwater Reservoir. *SPE Journal* **11**(4), 464-479.
- Journal, A. and Zhang T. [2006] The necessity of a multiple-point prior model. *Mathematical Geology*, **38**(5), 591-610.
- Lange, K., Cordua, K. S., Frydendall, J., Hansen, T. M. and Mosegaard, K. [2011] A frequency matching method for generation of a priori sample models from training images, *IAMG 2011*, Extended Abstract.
- Strebelle, S. [2002] Conditional simulation of complex geological structures using multiple-point geostatistics. *Mathematical Geology*, **34** (1), 1–22.

APPENDIX **F**

Paper 3

History Matching: towards Geologically Reasonable Models

Authors: Yulia Melnikova, Knud Skou Cordua and Klaus Mosegaard

Published in: Proceedings of EAGE conference, Dubai, UAE, 25-28 November 2012

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:

$$O(\mathbf{m}) = \frac{1}{2} \left\| d^{obs} - g(\mathbf{m}) \right\|_{C_d}^2 + f(\mathbf{m}, \text{TI}) \quad (1)$$

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:

$$h_{ij} = \frac{1}{(1 + Ad_{ij}^k)^p} \quad (2)$$

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 $f(\mathbf{m}, \text{TI})$ 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):

$$x_i = \ln \left(\frac{m_i - m^{low}}{m^{up} - m^i} \right) \quad (3)$$

Here $i = 1, \dots, n$, 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:

$$O^*(\mathbf{m}) = \left(\frac{\frac{1}{2} \|d^{obs} - g(\mathbf{m})\|_{C_d}^2 - h^*}{h^*} \right)^2 + \left(\frac{f(\mathbf{m}, \text{TI}) - f^*}{f^*} \right)^2 \quad (4)$$

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.

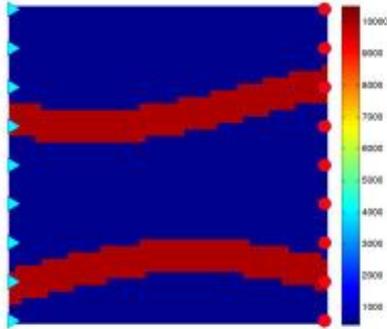


Figure 1
True permeability model, 49x49 pixels

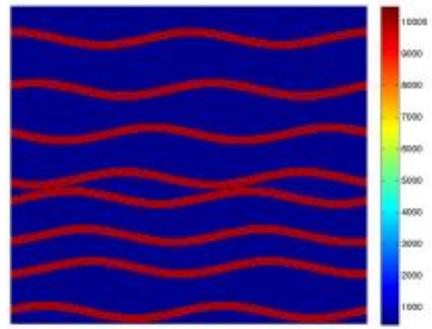


Figure 2
Training image, 200 x200 pixels

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)).

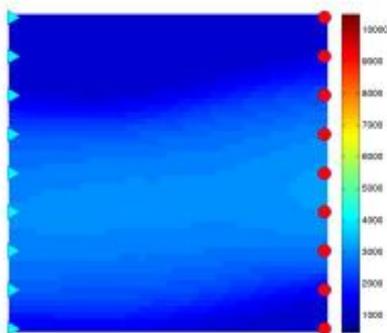


Figure 3 Initial permeability model

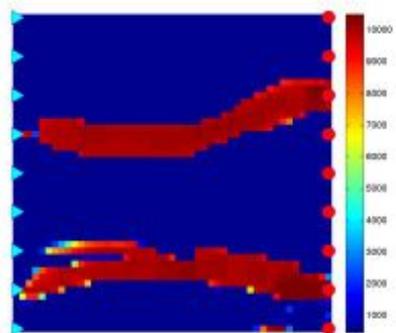


Figure 4 Solution, 97th iteration

Figure 5 demonstrates history matching for injector 4 and producer 3 (wells are numbered starting from the top).

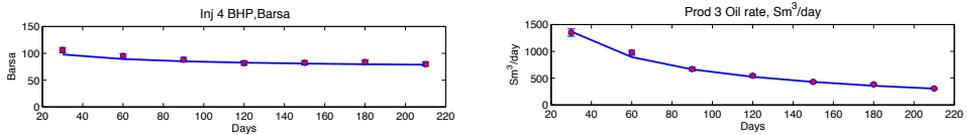


Figure 5 History matching: observed data (red circles) with error bars and solution response (blue line).

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.

Paper 4

History Matching with Geostatistical Prior: A Smooth Formulation

Authors: Yulia Melnikova, Katrine Lange, Andrea Zunino, Knud Skou Cordua and Klaus Mosegaard

Published in: Mathematics of Planet Earth

History matching with geostatistical prior: a smooth formulation

Yulia Melnikova, Katrine Lange, Andrea Zunino, Knud Skou Cordua and Klaus Mosegaard

Abstract We present a new method for solving the history matching problem by gradient-based optimization within a probabilistic framework. The focus is on minimizing the number of forward simulations and conserving geological realism of the solutions. Geological a priori information is taken into account by means of multipoint statistics borrowed from training images. Then production data and prior information are integrated into a single differentiable objective function, minimizer of which has a high posterior value. Solving the proposed optimization problem for an ensemble of different starting models, we obtain a set of solutions honouring both data and prior information.

Key words: history matching, multi-point statistics, gradient-based optimization

1 Introduction

History matching — inversion of reservoir production data for rock properties — is an ill-posed inverse problem with computationally expensive forward simulation. Highly non-linear relationship between data \mathbf{d}^{obs} and model parameters \mathbf{m} result in non-uniqueness of solutions. With the aid of geostatistical prior information (usually in the form of training images \mathbf{TI}) it has become possible to restrict the solution space drastically (Caers 2003; Jafarpour and Khodabakhshi 2011). The main challenge in history matching consists in minimizing the amount of forward simulations needed to achieve attractive solutions. In this work we present a new method for solving history matching problem using a probabilistic framework (Tarantola 2005), searching for solutions deterministically.

2 Methodology

Our approach consists in integrating production data and prior information into a single *differentiable* objective function, minimizer of which has a high posterior value:

$$\mathbf{m}^{\text{HighPosterior}} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_D}^2 + f^d(\mathbf{m}, \mathbf{TI}) \right\} \quad (1)$$

Y. Melnikova · K. Lange · A. Zunino · K.S. Cordua · K. Mosegaard
Department of Mathematical and Computational Geoscience (DTU Space) and CERE, Technical University of Denmark, Lyngby, Denmark; e-mail: yume@dtu.dk

where the first term is a conventional data misfit term (Tarantola 2005) and g represents forward simulation; the second term is the prior information misfit explained below (superscript d stands for differentiable). Solving Equation 1 by a gradient-based technique for an ensemble of starting models, we obtain a set of solutions with high-posterior value running a small number of forward simulations.

Lange et al (2012) define the prior misfit as the chi-square distance between multi-point statistics of the model and the training image, assuming both images to be discrete. Essentially, the statistics is the frequency distribution of multi-point patterns defined by a template. Contrary to Lange et al (2012), our formulation implies gradual change of the model parameters, therefore the prior misfit $f^d(\mathbf{m}, \mathbf{TI})$ should be defined for any continuous image. A continuous image can not be represented by the frequency distribution, however its differentiable approximation (a pseudo-histogram) can be computed.

Consider a set of size K of all unique discrete patterns observed in the training image. Then the pseudo-histogram of a continuous image is $H_j^{d,\text{image}} = \sum_{i=1}^{N^{\text{image}}} p_{ij}$, $j = 1, \dots, K$, where N^{image} is the number of patterns in the continuous image and p_{ij} is a measure of similarity between continuous and discrete patterns:

$$p_{ij} = \frac{1}{(1 + Ad_{ij}^k)^s}. \quad (2)$$

Here $d_{ij} = \|pat_i^{\text{image}} - pat_j^{\mathbf{TI}, \text{unique}}\|_2$ (Euclidean distance between pixel values of the corresponding patterns) and A , k , s are adjustable parameters. Per definition statistically similar images will have similar pseudo-histograms. Therefore we define the prior misfit as follows:

$$f^d(\mathbf{m}, \mathbf{TI}) = \sum_{j=1}^K \frac{(H_j^{d,\mathbf{TI}} - H_j^{d,\mathbf{m}})^2}{H_j^{d,\mathbf{TI}}}. \quad (3)$$

Use of the pseudo-histogram of the training image as a weight factor in Equation 3 results in proper reproduction of the pattern statistics.

3 Numerical example

We perform history matching on a 2D synthetic oil reservoir model. The goal is to obtain a set of permeability models having high posterior values. Others parameters, such as porosity, relative permeabilities and initial saturations, are assumed to be known. The reservoir model has 50 by 50 cells of size 10 by 10 meters. Figure 1a shows the true permeability and locations of injectors (down-triangles) and producers (up-triangles). Production data are generated by running a forward simulation with the true model and applying 5% Gaussian noise to the calculated water and oil rates. Physics of the flow (two-phase immiscible displacement) allows us to use few observations to perform history matching and spend less time computing sensitivities. We use only 2 measurements for each well (at 100 and 200 days), therefore 52 measurements in total. However, we show the full history to assure the quality of history matching.

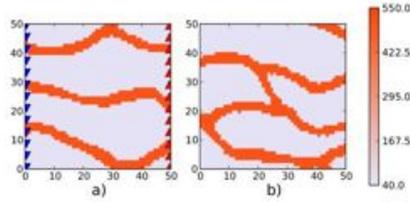


Fig. 1: a) True permeability model (channels - 500 mD, background - 50 mD); b) Training image (channels - 500 mD, background - 50 mD)

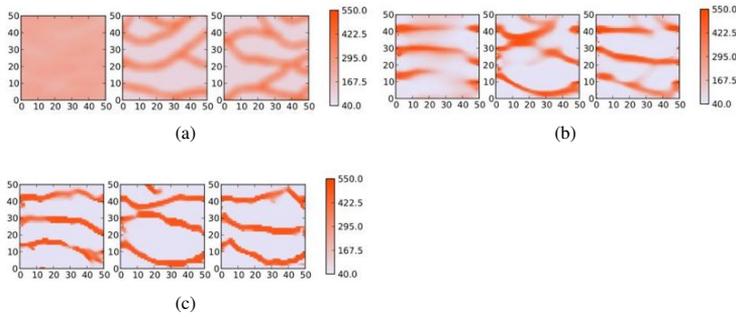


Fig. 2: a) Starting permeability models; b) Models after 30 iterations; c) Solutions for each of the starting models above

The prior information is given by a training image (Fig 1b), which is an upscaled part of the Strebelle image (Strebelle 2000). A square template of 6x6 pixels is used for collecting pattern statistics. Parameters A , k and s (Eq. 2) are set to 100, 2 and 2 respectively (empirically optimal values). Starting models (Fig 2a) are smoothed, upscaled parts of the Strebelle image (Strebelle 2000); after 30 iterations they are turned into models shown in Figure 2b (LBFGS optimization algorithm was used (Gao and Reynolds 2006)). Figure 2c shows final equally good solutions.

In all cases the pattern statistics of the training image is successfully reproduced, and the expected prior misfit value of 100 is achieved. Production data are resolved, since the data misfit value is everywhere $\approx N/2$, where N is the number of measurements (Oliver et al 2008). Data matching and convergence plots (for the first solution) are shown in Figure 3a and Figure 3b respectively. Naturally, the convergence properties of the algorithm are dependent on the initial guess and quality of the production and statistical data. Multiple solutions found in this example is a natural consequence of the fact that the history matching problem is a strongly un-

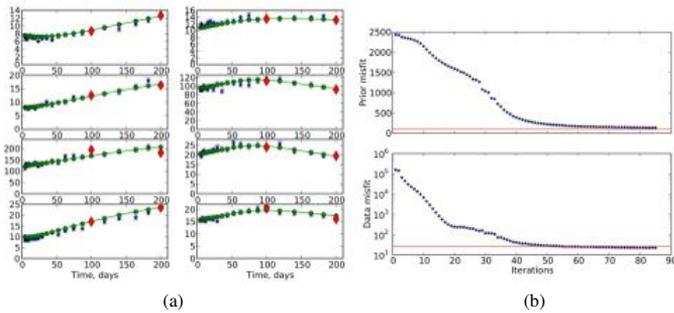


Fig. 3: a) History matching: water injection rates (left column) and oil production rates (right column) of some wells, m³/day; b) Convergence for prior and production data misfits. Expected values of the misfits are shown by red lines.

derdetermined problem. Thoroughly chosen initial guesses, obtained, for instance, from seismic data inversion, would be helpful in minimizing divergence of the solutions.

4 Discussion

We demonstrated how an ensemble of starting models can be gradually transformed into valuable solutions of the history matching problem. The suggested formulation has several advantages: 1) it guarantees prior-consistent solutions by including complex a priori information, 2) it allows using gradient-based optimization techniques, which save computational time, 3) it provides quantitative estimates of the data and prior information misfits and therefore allows us to distinguish between solutions as well as to choose the most reliable ones.

References

- Caers J (2003) History matching under training-image-based geological model constraints. *SPE J* 8:218–226
- Gao G, Reynolds AC (2006) An improved implementation of the lbfgs algorithm for automatic history matching. *SPE Journal* 11(1):5–17
- Jafarpour B, Khodabakhshi M (2011) A probability conditioning method (pcm) for nonlinear flow data integration into multipoint statistical facies simulation. *Math Geosci* 43:133–164
- 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. *Math Geosci* 44:783–803
- Oliver DS, Reynolds AC, Liu N (2008) *Petroleum reservoir characterization and history matching*. Cambridge University Press, New York.
- Strebelle S (2000) *Sequential simulation drawing structures from training images*. PhD thesis, Stanford University
- Tarantola A (2005) *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia

APPENDIX **H**

Paper 5

Discrete-facies History Matching in a Smooth Formulation

Authors: Yulia Melnikova, Andrea Zunino, Katrine Lange, Knud Skou Cordua and Klaus Mosegaard

Published in: Accepted to Mathematical Geosciences with revision

Mathematical Geosciences

Discrete-facies History Matching in a Smooth Formulation

--Manuscript Draft--

Manuscript Number:	MATG1075
Full Title:	Discrete-facies History Matching in a Smooth Formulation
Article Type:	Original Research
Keywords:	history matching; multiple-point statistics; optimization; inverse problems
Corresponding Author:	Yulia Melnikova Technical University of Denmark Kongens Lyngby, DENMARK
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Technical University of Denmark
Corresponding Author's Secondary Institution:	
First Author:	Yulia Melnikova
First Author Secondary Information:	
Order of Authors:	Yulia Melnikova Andrea Zunino Katrine Lange Knud Skou Cordua Klaus Mosegaard
Order of Authors Secondary Information:	
Abstract:	Solving the history matching problem, we focus on developing a probabilistic framework that integrates complex prior information and simultaneously allows us to sufficiently cut-down the number of forward simulations. The proposed smooth formulation combines the training-image-based prior with the production data into a single differentiable objective function, minimizer of which has a high posterior value. As the result, we obtain solutions, that honor the geological information and data, by applying gradient-based optimization. We propose a closed form expression for calculating the prior probabilities using theory of multinomial distributions, that allows us to rank the models in accordance with their relative posterior probabilities. yume@dtu.dk

Discrete-facies History Matching in a Smooth Formulation

Yulia Melnikova · Andrea Zunino ·

Katrine Lange · Knud Skou Cordua ·

Klaus Mosegaard

Received: date / Accepted: date

Abstract Solving the history matching problem, we focus on developing a probabilistic framework that integrates complex prior information and simultaneously allows us to sufficiently cut-down the number of forward simulations. The proposed smooth formulation combines the training-image-based prior with the production data into a single differentiable objective function, minimizer of which has a high posterior value. As the result, we obtain solutions, that honor the geological information and data, by applying gradient-based optimization. We propose a closed form expression for calculating the prior probabilities using theory of multinomial distributions, that allows us to rank the models in accordance with their relative posterior probabilities.

Keywords history matching · multiple-point statistics · optimization · inverse problems

Y. Melnikova · A. Zunino · K. S. Cordua · K. Mosegaard

Center for Energy Resources Engineering, Department of Mathematical and Computational Geoscience, National Space Institute, Technical University of Denmark, Richard Petersens Plads, Building 305, DK-2800, Kongens Lyngby, Denmark

E-mail: yume@dtu.dk

K. Lange

Banedanmark

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Introduction

History matching is a task of inferring knowledge about subsurface models of oil reservoirs from production data. History matching is a strongly underdetermined problem: having data in a limited number of wells, one needs to estimate rock properties in the whole reservoir model. This problem has infinitely many solutions, and in addition, most of them lack geological realism. Furthermore, the intensive computational work needed to simulate the data redoubles the complexity.

To address these challenges, we develop a probabilistic framework that combines complex a priori information and simultaneously aims at reducing number of forward simulations needed for finding solutions. We propose a smooth formulation of the inverse problem with discrete-facies prior defined by a multiple-point statistics model. This allows us to use gradient-based optimization methods to search for feasible models.

In probabilistic inverse problem theory (Tarantola 2005) the solution of an inverse problem is represented by its a posteriori probability density function (PDF). Each possible state in the model space is assigned a number — a posteriori probability density — which reflects how well the model honors the data and the a priori information (knowledge about the model parameters independent from the data). The a posteriori PDF of high-dimensional, underdetermined inverse problems, such as history matching, may feature isolated islands of significant probabilities and low probabilities everywhere else. Therefore, when the full description of the posterior PDF is not possible, the goal is to locate and explore islands of significant posterior probabilities.

One may explore the a posteriori PDF in several ways. Monte Carlo methods (Mosegaard and Tarantola 1995; Cordua et al 2012b) allow, in principle, sampling of the a posteriori PDF. However, for large scale non-linear inverse problems, there is a risk of detecting only a local island of significant posterior probability. In addition, sampling is not feasible for inverse problems with computationally expensive forward simulation, such as history matching. Other methods rely on

1 optimization (Caers and Hoffman 2006; Jafarpour and Khodabakhshi 2011) to
2 determine a collection of models that fit the data and the a priori information.
3
4 However, these methods fail to describe a posteriori variability of the models as
5 the weighting of prior information versus data information (likelihood) is not taken
6 into account.
7
8

9
10
11 Regardless of the chosen philosophy, most of the research community favor the
12 advanced prior information that helps to significantly shrink the solution space of
13 legal models (Caers 2003; Jafarpour and Khodabakhshi 2011; Hansen et al 2012).
14 For instance, the a priori information borrowed from a training image (Guardiano
15 and Srivastava 1993; Strebelle 2002) would permit only models of a specific con-
16 figuration defined by statistical properties of the image. Ideally, training images
17 reflect expert knowledge about geological phenomena (facies geometry, contrast in
18 rock properties, location of faults) and play a role of vital additional information,
19 drastically restricting the solution space (Hansen et al 2009).
20
21
22
23
24
25
26
27

28 Our strategy for exploring the a posteriori PDF, which is especially suitable
29 for inverse problems with expensive forward simulation (e.g. history matching), is
30 to obtain a set of models that have high posterior values, and rank the solutions
31 afterwards in accordance with their relative posterior probabilities. We integrate
32 complex a priori information represented by multiple-point statistics inferred from
33 a training image. One of the challenges here is to define a closed form expression
34 for the prior probability that, multiplied by the likelihood function, would provide
35 the a posteriori probability. It is not sufficient to perturb the model in consistency
36 with the training image until the dynamic data are matched as it is done in the
37 probability perturbation method (Caers and Hoffman 2006). As it was noticed
38 by Hansen et al (2012), in this method the fit to the prior information is not
39 quantified, so the method will spot models of maximum likelihood/nonzero prior,
40 not of the maximum posterior; the resulting model may resemble the training
41 image very poorly, and therefore may have a low posterior value.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Lange et al (2012) were the first who aimed at estimating prior probabilities
2 solving inverse problems with training images. The developed Frequency Matching
3 (FM) method is able to quantify the prior probability of a proposed model and
4 hence to iteratively guide it towards the high posterior solution. Specifically, Lange
5 et al (2012) solve a combinatorial optimization problem, perturbing the model in
6 a discrete manner until it explains both data and a priori information. In practice,
7 this requires many forward simulations and can be prohibitive for the history
8 matching problem.
9

10 While following the philosophy of the Frequency Matching method, we are
11 interested in minimizing the number of forward simulations needed to achieve a
12 model of a high posterior probability. Similar to the FM method, we minimize the
13 sum of the data and the prior misfits. However, a new smooth formulation of the
14 objective function allows us to apply gradient-based optimization and sufficiently
15 cut-down the number of reservoir simulations. After convergence the model has all
16 statistical properties of the training image and simultaneously fits the data. Having
17 several starting models, possibly very different, we are able to obtain different
18 solutions of the inverse problem and to detect places of high posterior probability.
19 In case of the history matching problem, starting models obtained from seismic
20 data interpretation probably would be of most practical use.
21

22 To our knowledge, gradient based techniques were first coupled with training
23 images in the work of Sarma et al (2008) by means of Kernel PCA. The authors
24 were the first who used Kernel Principal Component Analysis (PCA) for geolog-
25 ical model parametrization. The Kernel PCA generates differentiable (smooth)
26 realizations of the training image, maintaining its multiple-point statistics and,
27 as a result, reproducing geological structures. The differentiable formulation by
28 (Sarma et al 2008) allows the use of gradient-based methods; however the quality
29 of the solution in terms of consistency with the prior information is not estimated.
30 In this work we actually derive a closed form expression for the prior probability.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 This allows us to quantify the relative posterior probabilities of the solutions and
 2 therefore to assess their importance.
 3

4 This paper is organized as follows. In Sect. 2 we provide a smooth formulation
 5 of the inverse problem with training-image-defined prior. We introduce the concept
 6 of the smooth histogram and suggest a closed form expression for computing prior
 7 probabilities. In Sect. 3 we solve a two dimensional history matching problem and
 8 rank solutions in accordance with their relative posterior probabilities. Section 4
 9 summarizes our findings.
 10
 11
 12
 13
 14

15 **2 Methodology**

16 In this work we use a probabilistic formulation of the inverse problem, integrating
 17 complex a priori information with production history. The algorithm integrates dy-
 18 namic data and a priori information into a single differentiable objective function,
 19 minimizer of which has high-posterior value. Solving the optimization problem
 20 for an ensemble of starting models we obtain a set of solutions that honor both
 21 the observations and prior information. We start with a definition of the inverse
 22 problem.
 23
 24
 25
 26
 27
 28
 29
 30

31 **2.1 Inverse problem with training-image-defined prior**

32 Denoting the model parameters as \mathbf{m} , the non-linear forward operator as g and
 33 its response as \mathbf{d} we introduce the forward problem:
 34
 35
 36
 37
 38
 39

$$40 \quad \mathbf{d} = g(\mathbf{m}). \quad (1)$$

41 The inverse problem is defined then as the task of inferring the model parameters \mathbf{m}
 42 given the observed data \mathbf{d}^{obs} , the forward relation g and, if available, some (data
 43 independent) a priori information about model parameters. Addressing inverse
 44 problems, we employ a probabilistic approach (Tarantola 2005), where the solution
 45 is characterized by its a posteriori probability density function. The a posteriori
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65

PDF $\sigma(\mathbf{m})$ contains the combined information about the model parameters as provided by the a priori PDF $\rho(\mathbf{m})$ and the likelihood function $L(\mathbf{m})$:

$$\sigma(\mathbf{m}) = k \rho(\mathbf{m}) L(\mathbf{m}), \quad (2)$$

where k is a normalization constant. The likelihood function $L(\mathbf{m})$ shows how well the model \mathbf{m} fits the observations \mathbf{d}^{obs} :

$$L(\mathbf{m}) = c \exp\left(-\frac{1}{2} \|\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_{C_D}^2\right) \quad (3)$$

where c is a constant and C_D is the covariance matrix representing Gaussian uncertainties in the measurements.

Prior information is assumed to be obtained from a training image with discrete pixel (voxel) values, representing some subsurface property. In this case the expression for the a priori probability density function is actually known explicitly from the work of Lange et al (2012):

$$\rho(\mathbf{m}) = \text{const} \cdot \exp(-\alpha f(\mathbf{m}, \mathbf{TI})) \quad (4)$$

where the function $f(\mathbf{m}, \mathbf{TI})$ measures the similarity between the multiple-point statistics of the training image \mathbf{TI} and the model \mathbf{m} ; *const.* is the normalization constant, α is the problem-dependent weight factor. The statistics has the form of the frequency distribution of the observed patterns in the image. A pattern is a set of neighbouring pixels in the image of shape defined by the template T . Consider, for instance, a 2x2 square template applied to the binary image in Fig. 1a and the obtained histogram shown in Fig. 1c with dark-blue color (only non-zero counts out of possible 16 combinations are shown). Lange et al (2012) use the chi-square distance between the histogram of the training image and the histogram of the model to estimate their statistical similarity.

[Fig. 1 about here.]

Equation 3 and 4 immediately define the value of the posterior probability of a model (Eq. 2). Lange et al (2012) find the maximum a posteriori solution of the inverse problem minimizing the following sum of misfits:

$$\mathbf{m}^{MAP} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_D}^2 + \alpha f(\mathbf{m}, \mathbf{TI}) \right\} \quad (5)$$

The Frequency Matching method requires the search space to be discrete. Therefore Eq. 5 is a combinatorial optimization problem that typically requires running a large number of forward simulations. Lange et al (2012) used simulated annealing algorithm for the optimization ending up with several thousands of forward runs needed to achieve the solution. Aiming at minimizing the number of forward simulations (reservoir simulations) we suggest an alternative approach, where we replace categorical variables with continuous variables, thereby turning the combinatorial optimization problem into a continuous one and applying gradient-based search.

2.2 Approximating the frequency distribution

For clarity we use two-dimensional images in our explanation, while the approach is suitable for three dimensional problems as well. Assume that the prior information is represented by a categorical training image \mathbf{TI} . The goal is to gradually change a continuous starting model \mathbf{m} into a model $\mathbf{m}^{\text{HighPosterior}}$ with the high posterior value, i.e - one that honors both data and the prior information. We construct the following differentiable objective function:

$$\mathbf{m}^{\text{HighPosterior}} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_D}^2 + f^d(\mathbf{m}, \mathbf{TI}) \right\} \quad (6)$$

where superscript d stands for *differentiable*. Notice the absence of the weight factor α in comparison with Eq. 5. Constructing $f^d(\mathbf{m}, \mathbf{TI})$, we keep in mind the idea of representing and comparing images by their frequency distributions. However,

1
2 instead of computing the frequency distribution, we introduce its differentiable
3 approximation that can be computed for any continuous image as well.
4

5 Our notation is presented in Table 1.
6

7 [Table 1 about here.]
8

9
10 Pixel values in patterns are represented by real numbers. Notice that categorical
11 images are now treated as continuous. First we collect pattern statistics from the
12 **TI** and save its unique patterns as a database. Then for both **m** and **TI** their
13 pseudo-histograms are constructed. For **image** the pseudo-histogram $H_j^{d,\mathbf{image}}$ is
14 a vector of the length equal to the number of unique patterns in the **TI**. $H_j^{d,\mathbf{image}}$
15 reflects the “contribution” of all patterns found in **image** to $pat_j^{\mathbf{TI},\mathbf{unique}}$.
16
17
18
19

$$20 \quad H_j^{d,\mathbf{image}} = \sum_{i=1}^{N^{\mathbf{image}}} p_{ij}^{\mathbf{sim}} \quad (7)$$

21 where $p_{ij}^{\mathbf{sim}}$ defines the level of similarity between $pat_i^{\mathbf{image}}$ and $pat_j^{\mathbf{TI},\mathbf{unique}}$.
22

23 We define $p_{ij}^{\mathbf{sim}}$ such that it equals 1 when $pat_i^{\mathbf{image}}$ is pixel-wise equal to $pat_j^{\mathbf{TI},\mathbf{unique}}$.
24

25 A natural choice for $p_{ij}^{\mathbf{sim}}$ would be one based on the Euclidean distance between
26 pixel values of the corresponding patterns, defined, for instance, as below:
27
28
29

$$30 \quad p_{ij}^{\mathbf{sim}} = \frac{1}{(1 + A t_{ij}^k)^s} \quad (8)$$

31 where $t_{ij} = \|pat_i^{\mathbf{image}} - pat_j^{\mathbf{TI},\mathbf{unique}}\|_2$ and A, k, s are the user-defined param-
32 eters.
33
34

35 Notice the following property:
36
37
38

$$39 \quad p_{ij}^{\mathbf{sim}} = \begin{cases} 1 & t_{ij} = 0 \\ \in (0, 1) & t_{ij} \neq 0 \end{cases} \quad (9)$$

40 The smooth histogram computed for the discrete Image A (Fig 1a) is shown
41 in Fig. 1c by light-blue color. Figure 1b shows a continuous image, while in Fig.
42 1c one can see its histogram, defined in the smooth sense, depicted by the orange
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

color. Notice the small counts everywhere: indeed, according to Eq. 9, this image does not contain patterns sufficiently close to those observed in the training image.

The choice of A , k , s is very important: from one side, they define how well the pseudo-histogram approximates the true frequency distribution; from the other side, they are responsible for “smoothing” and, consequently, for the convergence properties. Figure 2 reflects how different values of k , s with fixed $A = 100$ influence the shape of the patterns similarity function (distance is normalized). Our empirical conclusion is that values $A = 100$, $k = 2$, $s = 2$ are optimal. Compare them (Fig 2) with the extreme case $A = 100$, $k = 1$, $s = 2$ where the majority of patterns have a close-to-zero contribution. These parameters are applicable after t_{ij} has been normalized on the quantity representing maximum possible Euclidean distance between the discrete patterns.

[Fig. 2 about here.]

2.3 Similarity function

Per definition statistically similar images will have similar pseudo-histograms.

Therefore we introduce the similarity function:

$$f^d(\mathbf{m}, \mathbf{TI}) = \frac{1}{2} \sum_{i=1}^{N^{\mathbf{TI}, unique}} \frac{(H_i^{d, \mathbf{m}} - H_i^{d, \mathbf{TI}})^2}{H_i^{d, \mathbf{TI}}} \quad (10)$$

Essentially, it is a weighted L2 norm, where the role of the weight parameter is played by the smooth histogram of the training image. The suggested measure favors patterns that are encountered less frequently in the training image and facilitates proper reproduction of the training image features. If number of patterns in the training image $N^{\mathbf{TI}}$ differs from the number of patterns in the model $N^{\mathbf{m}}$, we multiply $H_i^{d, \mathbf{TI}}$ by the following ratio:

$$r = \frac{N^{\mathbf{m}}}{N^{\mathbf{TI}}} \quad (11)$$

1 The use of the suggested similarity measure (Eq.10) can be validated through
 2 the following reasoning. Our idea consist in representing an image as an outcome
 3 of some multinomial experiment (see also Cordua et al (2012a)). Consider two
 4 categorical images: training and test. Assume, that a pattern in the test image is a
 5 multiple-point event that leads to the success for exactly one of K categories, where
 6 each category has a fixed probability of success p_i . By definition, each element H_i
 7 in the frequency distribution \mathbf{H} indicates the number of times the i^{th} category has
 8 appeared in N trials (number of patterns observed in the test image). Then the
 9 vector $\mathbf{H} = (H_1, \dots, H_K)$ follows the multinomial distribution with parameters N
 10 and \mathbf{p} , where $\mathbf{p} = (p_1, \dots, p_K)$

$$20 \quad P(\mathbf{H}) = P(H_1, \dots, H_K, N, p_1, \dots, p_K) = \frac{N!}{H_1! \dots H_K!} p_1^{H_1} \dots p_K^{H_K} \quad (12)$$

23 Let us assume that the test image is a realization of the random process that
 24 generated the training image. Then the vector of probabilities \mathbf{p} can be obtained
 25 from the frequency distribution of the training image $\mathbf{H}^{\mathbf{T}I}$: normalizing its entries
 26 on the total number of counts we obtain the probabilities of success. In other
 27 words, we assume that the histogram of the training image defines the theoretical
 28 distribution underlying the multinomial experiment. In general, the histogram of
 29 the training image is very sparse, therefore many categories of patterns will be
 30 assigned zero probabilities. It means that if a test image has a single pattern that
 31 is not encountered in the training image, its prior probability from Eq. 12 will be
 32 zero. It happens due to the insufficient prior information derived from the training
 33 image; it is very likely that many of the non-observed patterns have some non-zero
 34 probabilities to be observed. This problem is well-known in the field of the natural
 35 language processing (NLP): small vocabulary can imply zero probabilities of some
 36 words to exist. The NLP research community address the challenge with a fun-
 37 damental technique called “smoothing” (Chen and Goodman 1999). The common
 38 idea of smoothing algorithms lies in making prior distributions more uniform by
 39 adjusting low probabilites upward and high probabilities downward. Since there

is no information about the probabilities of the patterns not encountered in the training image, we assume them to be equal to ε . To make the sum of p_i equal to one, we subtract a small number γ from all non-zero bins of $\mathbf{H}^{\mathbf{TI}}$:

$$p_i = \begin{cases} \frac{H_i^{\mathbf{TI}} - \gamma}{N^{\mathbf{TI}}} & H_i^{\mathbf{TI}} > 0 \\ \varepsilon & H_i^{\mathbf{TI}} = 0 \end{cases} \quad (13)$$

where $\gamma = \varepsilon(K - N^{\mathbf{TI}, \text{unique}})N^{\mathbf{TI}}/N^{\mathbf{TI}, \text{unique}}$

This simple technique called absolute discounting is one of the many smoothing techniques, however, it is a topic of a separate research to define which smoothing methodology is the best for the training-image-based prior and we do not address it here.

After p_i having been defined in some manner, $P(\mathbf{H})$ can be computed through its logarithm:

$$\log(P(\mathbf{H})) = \log\left(\frac{N!}{H_1! \dots H_K!}\right) + \sum_{i=1}^K H_i \log(p_i) \quad (14)$$

We apply Stirling's approximation:

$$\log(n!) = n \log n - n + O(\log n) \quad (15)$$

Defining $I = \{i : H_i > 0\}$ we have:

$$\begin{aligned} \log\left(\frac{N!}{H_1! \dots H_K!}\right) &= \log(N!) - \sum_{i \in I} \log(H_i!) \approx N \log N - N - \\ &\sum_{i \in I} (H_i \log(H_i) - H_i) = N \log N - \sum_{i \in I} H_i \log(H_i) \end{aligned} \quad (16)$$

And finally,

$$\log(P(\mathbf{H})) \approx N \log N + \sum_{i \in I} H_i \log\left(\frac{p_i}{H_i}\right) = \sum_{i \in I} H_i \log\left(\frac{N p_i}{H_i}\right) \quad (17)$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Then

$$-\log(P(\mathbf{H})) \approx \sum_{i \in I} H_i \log\left(\frac{H_i}{Np_i}\right) \quad (18)$$

Substituting H_i with $Np_i + \varepsilon_i$ and applying Taylor expansion of the second order one arrives to the chi-square distance divided by two:

$$-\log(P(\mathbf{H})) \approx \frac{1}{2} \sum_{i \in I} \frac{(H_i - Np_i)^2}{Np_i} \quad (19)$$

Notice that Eq.19 justifies our choice of the similarity function (Eq. 10). Indeed, by minimizing expression 10 we minimize the value defined by Eq.19 as well. Further, if we denote $\mathbf{h} = \mathbf{H}/N$, Eq.17 is transformed:

$$\log(P(\mathbf{H})) \approx \sum_{i \in I} N h_i \log\left(\frac{p_i}{h_i}\right) = - \sum_{i \in I} N h_i \log\left(\frac{h_i}{p_i}\right) = -N D_{KL}(h||p) \quad (20)$$

where $D_{KL}(h||p)$ is the Kullback-Leibler divergence, a dissimilarity measure between two probability distributions h and p . In other words, it defines the information lost when the theory (training image) is used to approximate the observations (test image).

Having at hand a discrete image, one can compute its relative prior probability using Eq. 17. Moreover, it is also applicable to the continuous image at convergence, since the algorithm aims at finding solutions (images), pixel values of which are very close to the expected categorical values and therefore its patterns can be considered as a success in the multinomial experiment.

2.3.1 Generating prior realizations

Minimizing Eq. 10, we are able to generate a realization of the prior represented by a training image, given a starting guess. Consider a training image (Fig. 3), which is an upscaled part of a training image proposed by Strebelle (Strebelle 2000). We assume that it represents permeability of an oil reservoir with 500 mDarcy in channels and 10 mD in the background.

[Fig. 3 about here.]

To derive the multiple-point statistics, we used a square template of 6x6 pixels (optimal size according to the entropy approach suggested by Honarkhah (2011)). The training image has 789 unique 6x6 patterns, therefore the pseudo-histograms (Eq. 10) have 789 bins. Parameters A , k and s (Eq. 8) were set to the empirically optimal values: 100, 2, and 2. We solve the following optimization problem:

$$\mathbf{m}^{\text{HighPrior}} = \underset{\mathbf{m}}{\operatorname{argmin}} \left\{ f^d(\mathbf{m}, \mathbf{TI}) \right\} \quad (21)$$

Figure 4a shows three starting guesses: random, and upscaled smoothed parts of the aforementioned image of Strebelle (Strebelle 2000). Figure 4b shows state of the models after 20 iterations. And finally, Fig. 4c demonstrates the solutions obtained after 100 iterations. Since unconstrained optimization is used, the solutions have few outliers, however, the logarithmic transformation used in the optimization allows us to regulate the boundaries of pixel values. In this example the minimum possible value is 5 mD, and the maximum is 500 mD. The solutions clearly reproduce features of the training image. The value of prior misfit (Eq. 10) is close to 100.0.

[Fig. 4 about here.]

2.4 Optimization problem

It would be tempting to find the a high-posterior model by minimizing the objective function:

$$O(\mathbf{m}) = \frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_d}^2 + f^d(\mathbf{m}) \quad (22)$$

However, the two terms in this objective function have different dimensions and scales; this may lead to inconsistency in optimization. We overcome these difficulties transforming the current objective terms into dimensionless ones. For the current implementation we used the following expression (Oszczka 1978):

$$F_i^{trans}(x) = \frac{F_i(x) - F_i^*}{F_i^*}. \quad (23)$$

Here $F_i(x)$ is the i^{th} function to transform, and F_i^* is the target (desired) value of the objective function value. We denote the target value of the data misfit term as u^* , and from Oliver et al (2008) expect $u^* \approx N/2$, where N is the number of observations. The target value of the prior misfit f^* is not zero, since the training image and most of its realisations have slightly different histograms. However the order of magnitude of f^* , which corresponds to the well reproduced features of the training image, is the same and can be found empirically. It can be estimated by finding, for instance, the value of $f^d(m^*)$, where m^* is a realization of the training image. Alternatively the order of f^* can be found solving Eq. 21 for some starting model.

One of the easiest ways to combine objective functions into a single function is to use the weighted exponential sum (Marler and Arora 2004). We put equal weights on two misfit terms and the exponent equal to 2. This leads to the final expression for the objective function:

$$O^*(\mathbf{m}) = \left(\frac{\frac{1}{2} \|\mathbf{d}^{\text{obs}} - g(\mathbf{m})\|_{C_d}^2 - u^*}{u^*} \right)^2 + \left(\frac{f^d(\mathbf{m}, \mathbf{TI}) - f^*}{f^*} \right)^2 \quad (24)$$

Notice that the term with the largest difference between its current and target values gets higher priority. Essentially, u^* and f^* play roles of the weight, and the exact value is not needed to be known, the order is important. In practice, target values can be set below the desired values to provide faster convergence.

To be able to use unconstrained optimization in case of non-negative model parameters (such as permeability), we applied the logarithmic scaling of the parameters (Gao and Reynolds 2006):

$$x_i = \log \left(\frac{m_i - m^{low}}{m^{up} - m_i} \right). \quad (25)$$

1 Here $i = 1, \dots, n$, where n is the number of pixels in the test image m , m^{low} and m^{up}
2 are the lower and upper scaling boundaries, respectively, of the parameters. The
3 log-transform does not allow extreme values of the model parameters and makes
4 the algorithm perform in a more robust way. Additionally, the applied scaling of
5 model parameters results in better convergence properties of the algorithm (Gao
6 and Reynolds 2006).
7

8 For solving (24) we suggest using quasi-Newton methods that are known to
9 be efficient for history matching problems (Oliver et al. 2008). In this work we
10 employed the unconstrained implementation of the L-BFGS method (Zhu et al.
11 1997). The algorithm requires calculation of the objective function value and its
12 gradient (the Hessian needed for the search direction is evaluated by approxima-
13 tion (Nocedal and Wright 2006). The gradient of the data misfit term is calculated
14 by the adjoint method implemented in the reservoir simulator Eclipse (Schlum-
15 berger GeoQuest 2009). The gradient of the prior term is computed analytically.
16 The algorithm is stopped when the values of the objective terms in the optimiza-
17 tion problem (24) approach their target values. The computational efficiency of
18 the algorithm decreases with increase of the number of categories in the training
19 image or the template size, since larger number of Euclidean distances is to be
20 calculated.
21

22 **3 History matching**

23 We perform history matching on a two-dimensional synthetic oil reservoir, aiming
24 at estimating its permeability field. All other parameters, such as porosity, relative
25 permeabilities and initial saturations are assumed to be known. To investigate non-
26 uniqueness of the solutions we solve Eq. 24 for a set of starting models. Table 2
27 lists some parameters of the reservoir model.
28

29 [Table 2 about here.]

30 Figure 5a shows the true permeability field that features sand channels of 500
31 mD and background shale of 10 mD; 13 injectors are marked by triangles, and
32

13 producers by circles respectively. All wells work at the bottom hole pressure control: 300 Barsa for the injectors and 50 Barsa for the producers. Production data are generated by running a forward simulation with the true permeability model and adding 5 % of Gaussian noise. Physics of the flow (steady two-phase immiscible displacement) allows us to use few observations and not to lose in history matching accuracy. We choose just two measurements (at 100 and 200 days) per well, 52 measurements in total (we measure water rate in injectors and oil rate in producers). This approach results in faster performance, since much less time is required to compute sensitivities. However, we show full history to assure the quality of history matching.

[Fig. 5 about here.]

Prior information is given by the training image in Fig. 3. We used the same parameters as in Sect. 2.3.1 to derive multiple-point statistics and construct the objective function. The ensemble of starting guesses (Fig. 6a) is presented by randomly chosen parts of a smoothed and upscaled version of the training image proposed by Strebelle (Strebelle 2000).

Solving Eq. 24 we put target values of u^* and f^* as 10.00 and 25.00 to assure the convergence of the algorithm to the desired values of the misfits. For the data misfit we expect a value close to $\approx N/2$ where N is the number of measurements (Oliver et al 2008) and for the prior close to 1.e+2. On average the algorithm converges in 100 iterations; its performance depends on the closeness of the initial guess to the solution. Fig. 6b demonstrates the transformation of the models after 50 iterations: most of the original channels are blurred and new ones are being constructed. Fig. 6c shows models at the 150th iteration. The algorithm successfully reproduced high-contrast channels of the expected continuity and width. Production data assured correct location of channels, and in many cases they are very close to the true model. Naturally, since the data sensitivity decreases with the increase of the distance from a well, the location of channels is very well defined on the sides of the model in the vicinity of wells, while in the middle we observe

1
2 some deviation from the true model. This example clearly demonstrates the con-
3 sequences of the underdetermined inverse problem: existence of many solutions
4 satisfying the available information.
5

6 Figure 7a shows history matching for the first solution: injection rates of the
7 first four injectors and production rates of the first four producers (counting from
8 top). Convergence plot for the prior and the data misfit is shown in Fig. 7b (notice
9 log-scale for the data misfit term). Red lines mark the desired values of the misfits.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

[Fig. 6 about here.]

[Fig. 7 about here.]

Finally, we are able to distinguish between the solutions (Fig. 6c) by calculating for each of them the value of the relative posterior probability derived from Eq. 2 and Eq. 3:

$$\log(\sigma(\mathbf{m})/(k * c)) = \log(\rho(\mathbf{m})) - \frac{1}{2} \|g(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_{C_D}^2 \quad (26)$$

where $\log(\rho(\mathbf{m}))$ is defined by Eq.17. We chose $\gamma = 0.1$. Table 3 lists the results (numeration of the models starts from top).

[Table 3 about here.]

For the comparison, in the last row we give the value calculated for the true model (Fig. 5a). We can conclude that models 5, 8 and 9 are the most preferable within this ensemble, while model 3 is the most inferior.

4 Conclusions

We presented an efficient method for solving the history matching problem with the aid of gradient optimization integrating complex training- image-based prior information into the solution. History matching is a severely undetermined inverse problem and existence of multiple solutions is a direct (and unfortunate) consequence of this property. However, production data contain valuable information about rock properties, such as porosity and permeability. Inversion of them is necessary for construction of reservoir models that can be used in prediction. Geological information, if available, can drastically decrease the solution space.

One way of applying the methodology is to explore the solution space. Since we are able to start from any smooth model in many cases we can detect solutions that would have high posterior values and look very different, due to the fact that they would belong to the different islands of high probability. Quantification of

1 the relative posterior probabilities allows us to rank solutions and choose the most
2 reliable ones.

3
4 The algorithm needs a starting guess, and, clearly as in any gradient-based op-
5 timization, the convergence properties would depend on it. In the history matching
6 problem, the choice of the starting guess is particularly important. The sensitivity
7 of the production data with respect to the rock properties decreases non-linearly
8 with the distance from wells. Therefore it is hard to invert for model parameters
9 in the areas with the poor well coverage. The situation can be greatly simplified if
10 one would integrate seismic data, or at least, would use the results of the seismic
11 inversion as the starting guesses. Indeed, having in general high resolution, the
12 geophysical data may serve as a powerful constraining tool in the inversion. This
13 is a topic of our future research.

14
15
16
17
18
19
20
21
22
23 **Acknowledgements** The present work was sponsored by the Danish Council for Independent
24 Research - Technology and Production Sciences (FTP grant no. 274-09-0332) and DONG
25 Energy.

26 27 28 29 30 31 32 **References**

- 33
34
35 Caers J (2003) History matching under training-image-based geological model con-
36 straints. SPE J 8:218–226
37
38 Caers J, Hoffman T (2006) The probability perturbation method: a new look at
39 bayesian inverse modeling. Math Geol 38:81–100
40
41 Chen SF, Goodman J (1999) An empirical study of smoothing techniques for
42 language modeling. Comput Speech Lang 13:359–394
43
44 Cordua KS, Hansen TM, Lange K, Frydendall J, Mosegaard K (2012a) Improving
45 multiple-point-based a priori models for inverse problems by combining sequen-
46 tial simulation with the frequency matching method. In: 82th Ann. Meet. Soc.
47 Expl. Geophys
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

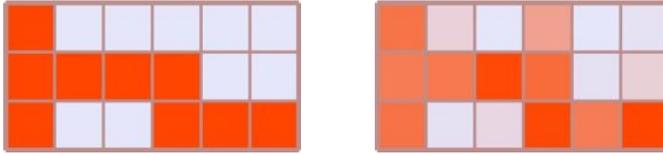
- 1 Cordua KS, Hansen TM, Mosegaard K (2012b) Monte carlo full waveform inversion
2 of crosshole gpr data using multiple-point geostatistical a priori information.
3 *Geophysics* 77:H19–H31
- 4
5
6 Gao G, Reynolds AC (2006) An improved implementation of the lbfgs algorithm
7 for automatic history matching. *SPE Journal* 11(1):5–17
- 8
9
10 Guardiano F, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate
11 moments. In: Soares, A. (ed.) *Geostatistics Troia*, Kluwer Academic, vol 1
- 12
13 Hansen TM, Mosegaard K, Cordua KS (2009) Reducing complexity of inverse
14 problems using geostatistical priors. In: *Proceedings of IAMG 09*
- 15
16 Hansen TM, Cordua KS, Mosegaard K (2012) Inverse problems with non-trivial
17 priors: efficient solution through sequential gibbs sampling. *Comput Geosci*
18 16:593–611
- 19
20
21 Honarkhah M (2011) Stochastic simulation of patterns using distance-based pat-
22 tern modeling. PhD thesis, Stanford University
- 23
24 Jafarpour B, Khodabakhshi M (2011) A probability conditioning method (pcm)
25 for nonlinear flow data integration into multipoint statistical facies simulation.
26 *Math Geosci* 43:133–164
- 27
28
29 Lange K, Frydendall J, Cordua KS, Hansen TM, Melnikova Y, Mosegaard K (2012)
30 A frequency matching method: solving inverse problems by use of geologically
31 realistic prior information. *Math Geosci* 44:783–803
- 32
33
34 Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for
35 engineering. *Struct Multidisc Optim* 26:369–395
- 36
37
38 Mosegaard K, Tarantola A (1995) Monte carlo sampling of solutions to inverse
39 problems. *J Geophys Res* 100:12,431–12,447
- 40
41
42 Nocedal J, Wright SJ (2006) *Numerical optimization*. Springer
- 43
44 Oliver DS, Reynolds AC, Liu N (2008) *Petroleum reservoir characterization and*
45 *history matching*. Cambridge University Press, New York.
- 46
47 Osyczka A (1978) An approach to multicriterion optimization problems for engi-
48 neering design. *Comput Meth Appl Mech Eng* 15:309–333
- 49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

-
- 1
2 Sarma P, Durlofsky LJ, Aziz K (2008) Kernel principal component analysis for
3 efficient, differentiable parameterization of multipoint geostatistics. *Math Geosci*
4 40:3–32
5
6 Schlumberger GeoQuest (2009) ECLIPSE reservoir simulator. Technical descrip-
7 tion
8
9 Strebelle S (2000) Sequential simulation drawing structures from training images.
10 PhD thesis, Stanford University
11
12 Strebelle S (2002) Conditional simulation of complex geological structures using
13 multiple-point statistics. *Math Geol* 34(1):1–20
14
15 Tarantola A (2005) Inverse problem theory and methods for model parameter
16 estimation. Society for Industrial and Applied Mathematics, Philadelphia
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

List of Figures

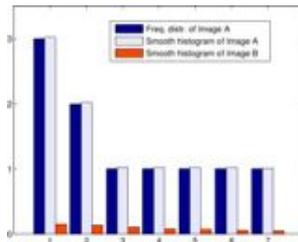
1	Frequency distribution and its approximation	22
2	Patterns similarity function	23
3	Training image	24
4	Generating prior realizations	25
5	True model of permeability	26
6	(a) Starting models, (b) Models after 50 iterations, (c) Models after 150 iterations	27
7	a) History matching for the first solution; b) Convergence of the prior and data misfits	28

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



(a) Discrete Image A

(b) Continuous image B



(c) Frequency distribution of Image A; Smooth histogram of Image A; Smooth histogram of Image B; 2x2 template applied

Fig. 1: Frequency distribution and its approximation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

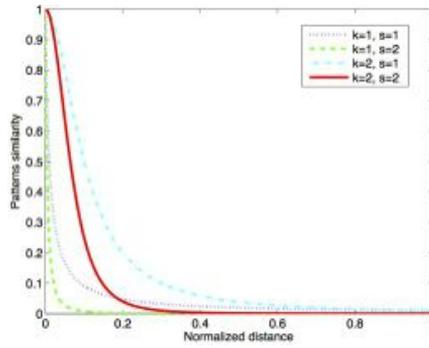


Fig. 2: Patterns similarity function

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

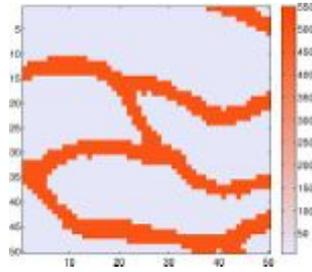


Fig. 3: Training image

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

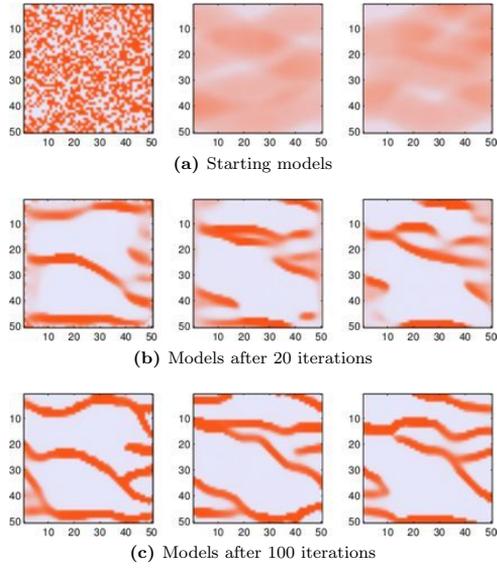


Fig. 4: Generating prior realizations

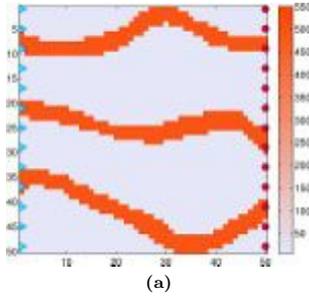


Fig. 5: True model of permeability

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

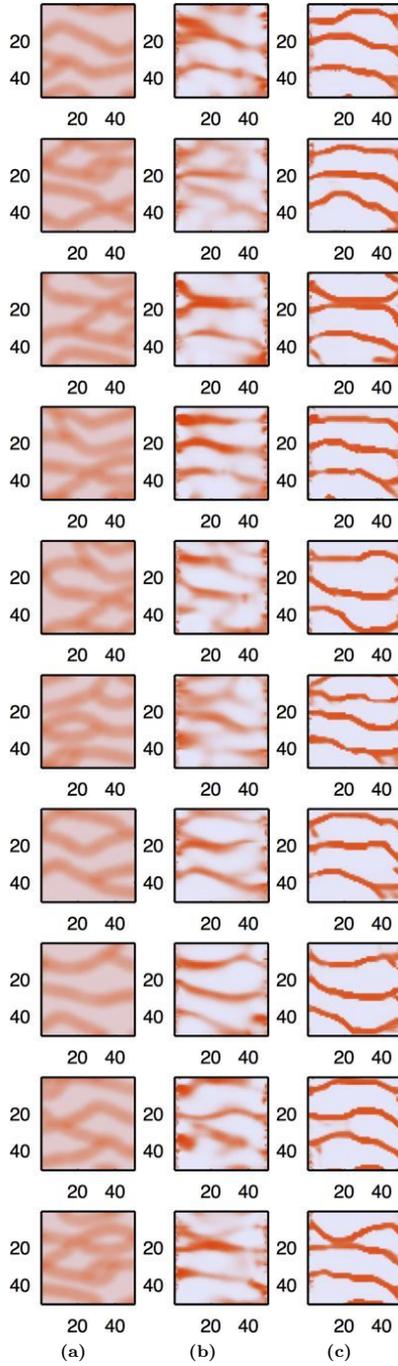


Fig. 6: (a) Starting models, (b) Models after 50 iterations, (c) Models after 150 iterations

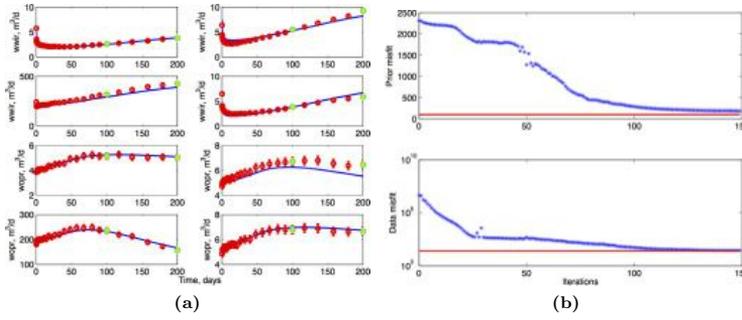


Fig. 7: a) History matching for the first solution; b) Convergence of the prior and data misfits

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

List of Tables

1	Notation	30
2	Reservoir model parameters	31
3	Posterior ranking of the solutions	32

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1: Notation

Notation	Description
TI	training image, categorical
m	model (test image), can contain continuous values
image	image (training or test)
T	scanning template
$H^{d,\mathbf{image}}$	pseudo-histogram of image
$pat_j^{\mathbf{TI},unique}$	j^{th} unique pattern in TI
$N^{\mathbf{image}}$	number of patterns in image
$pat_i^{\mathbf{image}}$	i^{th} pattern of pixels from image .

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 2: Reservoir model parameters

Model size	50x50 cells
Cell size	10x10 m
Initial water saturation	0.0
Porosity	0.3 (constant everywhere)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 3: Posterior ranking of the solutions

Model N	$\log(\rho(\mathbf{m})) - \frac{1}{2}\ g(\mathbf{m}) - \mathbf{d}^{\text{obs}}\ _{C_D}^2$
1	-8122.0324
2	-8134.6031
3	-10383.1467
4	-7860.2211
5	-6568.7915
6	-8900.5525
7	-9781.7611
8	-7107.3847
9	-6734.4299
10	-7608.2761
11	-7713.9272

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

APPENDIX



Paper 6

Reservoir Modeling Combining Geostatistics with
Markov Chain Monte Carlo Inversion

Authors: Andrea Zunino, Katrine Lange, Yulia Melnikova, Thomas Mejer
Hansen and Klaus Mosegaard

Published in: Mathematics of Planet Earth

Reservoir modeling combining geostatistics with Markov chain Monte Carlo inversion

Andrea Zunino, Katrine Lange, Yulia Melnikova, Thomas Mejer Hansen and Klaus Mosegaard

Abstract We present a study on the inversion of seismic reflection data generated from a synthetic reservoir model. Our aim is to invert directly for rock facies and porosity of the target reservoir zone. We solve this inverse problem using a Markov chain Monte Carlo (McMC) method to handle the nonlinear, multi-step forward model (rock physics and seismology) and to provide realistic estimates of uncertainties. To generate realistic models which represent samples of the prior distribution, and to overcome the high computational demand, we reduce the search space utilizing an algorithm drawn from geostatistics. The geostatistical algorithm learns the multiple-point statistics from prototype models, then generates proposal models which are tested by a Metropolis sampler. The solution of the inverse problem is finally represented by a collection of reservoir models in terms of facies and porosity, which constitute samples of the posterior distribution.

Key words: monte carlo, inversion, reservoir modeling, seismic reflection, rock physics

1 Introduction

Reservoir modeling conditioned by recorded seismic reflection data is the most prominent geophysical technique to investigate the unknown properties of the subsurface. However, even if seismology produces good quality tomographic images, it still remains challenging to obtain a good picture of some particular properties such as porosity or permeability that are of most interest for oil and gas exploration. The link between elastic parameters and such properties lies in the complex relationships between, among others, intrinsic properties of rocks, mineralogy, and interaction with fluids which are usually described by a rock physics model (Mavko et al, 2003). Since these relationships are usually nonlinear and affected by uncertainty, it is difficult to invert seismic data directly for, e.g., porosity employing the standard optimization approaches because they generally rely on linearised models and simple scaling laws. Here we propose an approach based on a Markov chain Monte Carlo (McMC) technique which is able to combine rock physics modeling and reflection seismology to invert for porosity and facies of the subsurface. It takes

A. Zunino · K. Lange · Y. Melnikova · T. Mejer Hansen · K. Mosegaard
Department of Mathematical and Computational Geoscience (DTU Space) and CERE, Technical University of Denmark, Lyngby, Denmark; e-mail: anzu@dtu.dk

into account the nonlinearities deriving from the rock physics model and moreover it provides an estimation of uncertainties on the unknown properties. Similar approaches have been studied before, see e.g., González et al (2008); Bosch et al (2010); Rimstad and Omre (2010).

2 Overview of the Markov chain Monte Carlo inverse method

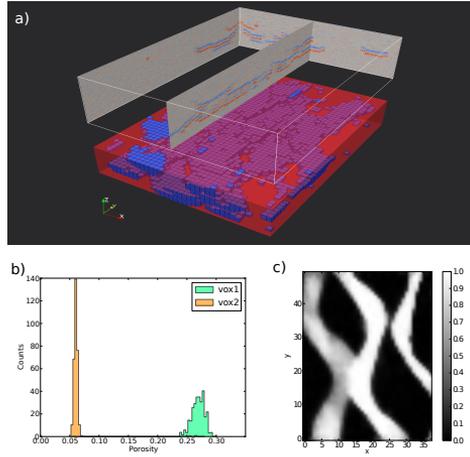
We follow a probabilistic approach, in which all information is represented by probabilities, as described in Tarantola (2005), where the inverse problem consists in performing an indirect measurement of unobservable parameters of the subsurface given some measured quantities on the surface of the Earth. The solution to the inverse problem is the posterior distribution, a combination of the prior and likelihood functions describing all possible models and relative probabilities.

Our aim is then to explore the model space in order to obtain a collection of models which all fit the measured data and are consistent with the a priori information. Moreover we are interested in estimating the uncertainty on unknown model parameters. Markov chain Monte Carlo algorithms represent a natural choice to fulfill these requirements, so we construct a multi-step algorithm capable of sampling the posterior distribution. The ingredients necessary to sample solutions to this inverse problem are essentially two (Mosegaard and Tarantola, 1995): I) an algorithm generating samples from a proposal distribution according to the available prior information and II) a sampler of the likelihood function. The prior geological information is represented by one or multiple training images which supply the necessary information about geological patterns to the algorithm. The posterior distribution is finally sampled employing the extended Metropolis algorithm (Metropolis et al, 1953; Mosegaard and Tarantola, 1995) based on the degree of fit between measured and calculated seismograms. We consider Gaussian uncertainties and hence we utilize an L_2 -norm for the misfit function.

Importance of informed priors - Geostatistics

One difficulty arising in high-dimensional space sampling is that a tremendous computational effort is needed to properly sample the posterior distribution. The huge size of model space, in fact, hampers the adoption of this kind of methodology in several cases. However, the use of proper informed priors can significantly improve the situation, reducing drastically the size of the model space to be sampled. This is obtained by employing an algorithm which generates models adhering to the prior knowledge so that only plausible models are taken into account in the sampling process. One recently introduced technique consists in generating realizations of a model exploiting the multiple-point statistics contained in prototype models. Specifically, the sequential Gibbs sampling method (see Hansen et al, 2012, and references therein) uses a sequential simulation approach where the algorithm learns the statistics from a training image which is scanned searching for recurring patterns. In principle, to increase the number of patterns, multiple training images may be used. A randomly selected hyper-rectangular volume of the model is then chosen to be re-simulated at each iteration of the Markov chain to propose a new model, where voxels are re-computed using sequential simulation conditioned on the rest of voxels (Strebelle, 2002).

Fig. 1 a) An example of a two-facies reservoir model from the collection of solutions with some slices through the volume of observed seismograms plotted on top. b) Histogram of porosity for two voxels, one located at $(x, y, z) = (1500, 3500, 20)$ m and the other at $(1000, 1000, 48)$ m. c) Probability of having sand (and hence a channel) on a 2D slice of the model at $z = 40$ m.



3 Numerical experiments

The target of our study is a synthetic reservoir model derived (but modified) from the Stanford VI-E model (Castro et al, 2005). It consists of a 3D arrangement of $38 \times 50 \times 20$ voxels with size of 100, 100 and 4m each respectively. Each voxel is parameterised with facies and porosity as the unknown parameters. Using the reservoir model derived from the Stanford VI-E model we constructed some “synthetic observations” by computing the seismograms to be inverted. In our case the forward model calculation consists of several steps. The first is the computation of the elastic properties from the facies and porosity of the subsoil. Then we compute the synthetic seismograms using a convolution approach.

The target zone of the reservoir is constituted by two facies, one representing sand (channel in a fluvial deposition system and oil-saturated) and the other representing shale (floodplain and brine-saturated). We assume the mineralogy to be known and describe it as consisting of four minerals (clay, quartz, feldspar, rock fragments) with known volume fraction in each facies but unknown porosity. The link between porosity and other petrophysical properties with the elastic moduli of the bulk rock for sand facies is modeled using the constant cement model (Dvorkin et al, 1994) and the usual formula for isotropic V_p . An empirical law from Gardner et al (1974) is used instead to compute V_p for shale facies.

Seismic modeling is carried out in the framework of the acoustic approximation, where the basic ingredients are the P-wave velocity and the density model. The seismic data are “recorded” at the surface on top of each pixel column as a zero-offset section. This in reality can correspond to data recorded at different source-receiver offset that have been processed such that they represent an equivalent zero-offset section which is easier to interpret. The wavelet is constructed from a Ricker

function with 50Hz peak frequency and is assumed to be known in the inversion process.

4 Results and discussion

We ran $2 \cdot 10^6$ iterations, obtaining about $7 \cdot 10^5$ models, of which only one every 10^2 was retained to ensure independence of samples. Fig. 1a shows one particular model from the solutions. We ended up with a collection of models representing samples of the posterior distribution which can be used to estimate subsurface properties and their relative probabilities/uncertainties. The solutions are used as a database that can be queried to obtain information on several different aspects since it represents the complete solution of the inverse problem. Here we show two examples of the kind of information which can be retrieved from the collection of models. The first is to compute the value of porosity at two different locations, obtaining histograms of possible values (Fig. 1b). The histogram tells us which range of values is most probable and, moreover, gives us an estimation of the uncertainty. The two histograms show a different behavior, one having a more pronounced peak, reflecting the different degree of resolving power. The second example is a map of the probability of having the sand facies on a slice of the 3D model at $z = 40\text{m}$ (Fig. 1c). The continuity of structures depicted in Fig. 1c is due to the prior information deriving from the geostatistical algorithm which takes into account the spatial continuity present in the training image. This example shows how it is possible to retrieve more sophisticated information from the database of solutions that can result very useful for real problems applications. Again, the uncertainty, clearly imaged in this probability plot, is an integral part of the answer we were searching for.

References

- Bosch M, Mukerji T, Gonzalez E (2010) Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics* 75(5):75A165–75A176
- Castro S, Caers J, Mukerji T (2005) The Stanford VI reservoir: 118th annual report. Tech. rep., Stanford Center for Reservoir Forecasting, Stanford University, Palo Alto, CA
- Dvorkin J, Nur A, Yin H (1994) Effective properties of cemented granular materials. *Mechanics of Materials* 18(4):351 – 366
- Gardner G, Gardner L, Gregory A (1974) Formation velocity and density – the diagnostic basics for stratigraphic traps. *Geophysics* 39(6):770–780
- González E, Mukerji T, Mavko G (2008) Seismic inversion combining rock physics and multiple-point geostatistics. *Geophysics* 73(1):R11–R21
- Hansen T, Cordua K, Mosegaard K (2012) Inverse problems with non-trivial priors: efficient solution through sequential gibbs sampling. *Computational Geosciences* 16:593–611
- Mavko G, Mukerji T, Dvorkin J (2003) *The Rock Physics Handbook*. Cambridge University Press
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6):1087–1092
- Mosegaard K, Tarantola A (1995) Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research* 100(B7):12,431–12,447
- Rimstad K, Omre H (2010) Impact of rock-physics depth trends and markov random fields on hierarchical bayesian lithology/fluid prediction. *Geophysics* 75(4):R93–R108
- 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 Model Parameter Estimation*. SIAM