HAL
open science

# Method for Stochastic Inverse Modeling of Fault Geometry and Connectivity Using Flow Data

Nicolas Cherpeau, Guillaume Caumon, Jef Caers, Bruno Lévy

## To cite this version:

Nicolas Cherpeau, Guillaume Caumon, Jef Caers, Bruno Lévy. Method for Stochastic Inverse Modeling of Fault Geometry and Connectivity Using Flow Data. Mathematical Geosciences, 2012, 44 (2), pp.147 - 168. 10.1007/s11004-012-9389-2 . hal-03029492

**HAL Id: hal-03029492**

**https://hal.univ-lorraine.fr/hal-03029492v1**

Submitted on 28 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Method for stochastic inverse modeling of fault geometry and connectivity using flow data *

Nicolas Cherpeau [1,2,3], Guillaume Caumon [3], Jef Caers [4] and Bruno Lévy [5]

*Published in *Mathematical Geosciences*, 44:147–168, 2012. ISSN 1874-8961. doi : 10.1007/s11004-012-9389-2

[1]Corresponding author

[2]e-mail: nicolas.cherpeau@gocad.org

[3]Centre de Recherches Pétrographiques et Géochimiques - Université Lorraine, École Nationale Supérieure de Géologie, Rue du doyen Marcel Roubault, 54501 Vandoeuvre-lès-Nancy, France

[4]Department of Energy Resources Engineering, Stanford university, Stanford, CA 94305, USA

[5]Centre INRIA Nancy Grand-Est, Campus scientifique, 615 rue du Jardin Botanique, 54600 Villers les Nancy, France

**Abstract**

This paper focuses on fault-related uncertainties in the subsurface, which can significantly affect the numerical simulation of physical processes. Our goal is to use dynamic data and process-based simulation to update structural uncertainty in a Bayesian inverse approach. We propose a stochastic fault model where the number and features of faults are made variable. In particular, this model samples uncertainties about connectivity between the faults. The stochastic 3D fault model is integrated within a stochastic inversion scheme in order to reduce uncertainties about fault characteristics and fault zone layout, by minimizing the mismatch between observed and simulated data.

The stochastic fault model uses a priori information such as fault orientation, location, size and sinuosity, to sample both geometrical and topological uncertainties with realistic fault descriptions. Each fault object is parameterized by the random vector used to simulate fault features. Then, during inversion, the random vector of the current model is stochastically perturbed, producing a new parameter vector used as input by the stochastic fault model to produce a new model. Even if the topology varies from one model to another, the algorithm produces correlated models so that their flow responses evolve quite smoothly.

The methodology is applicable in general and illustrated on a synthetic two-phase flow example. A first set of models is generated to sample the prior uncertainty space. Then, models minimizing reference water-saturation data misfit are used as seeds to generate continuous Markov chains of models with discrete states. Posterior models reduce uncertainties about fault position, while the topology varies from one model to another. A second example highlights the interest of the parameterization when interpreted data is available, by perturbing geological scenarios and falsifying those that do not match two-phase flow observations.

KEYWORDS: structural modeling; uncertainty; topology; inverse modeling

# 1 Introduction

In subsurface studies, uncertainties about faults and fault network primarily stem from the lack or poor quality of observation data. Due to the inherent uncertainties and possible human bias during interpretation of geological data (Bond et al., 2007), model predictions seldom match physical observations such as reservoir pressure, seismograph measurements or gravity anomalies. One may want to use such data to infer some characteristics of the model, which is known as the inverse problem.

In this paper, we propose a general method to reduce structural uncertainty about fault networks. The method generates models with varying number of faults, fault geometry and fault connectivity, i.e. how faults truncate each others. The method accounts for large to small scale faults (km to decameter scale), for which uncertainties may be large due to the lack of data (incomplete sampling or sensing), imaging ambiguities (e.g. in potential field and seismic methods), limited resolution and positioning errors (borehole trajectory, time-to-depth conversion, etc.). Therefore, in the context of hydrocarbon reservoir management, our method is applicable at several stages :

- In exploration of "green fields", in which a limited amount of data is generally available (a set of seismic lines, a few borehole data). In this case, we focus on large-scale, kilometric faults or larger.

- In "brown field" management, in which 3D seismic data is generally available but only provides approximate view on structures because of

resolution, acquisition or processing issues. In this case, the method is applicable to aid the interpretation of large to mid-scale faults in poorly imaged areas or to simulate sub-seismic faults following geological priors.

In such uncertain situations, the method mainly relies on prior information about faults, which may come from the interpretation of available data, structural geological concepts, regional context or reservoir analogs.

We apply this method to the inversion of reservoir or hydrogeologic flow data. Indeed, faults impact fluid flow in different ways:

1. Individually, faults have a direct impact on the permeability field due to fracturing, alteration, cementation or shale smearing nearby the fault. Uncertainties about fault rock hydraulic properties in flow models are generally high and may have a significant impact on fluid flow (Manzocchi et al., 1999).

2. At the reservoir scale, fault zone complexity and spatial layout directly impact fluid flow by creating connected flow conduits or flow barriers isolating fault blocks, depending on the nature of the fault rock. In this paper, we focus on uncertainties about fault layout and geometry including fault connectivity uncertainties.

3. The displacement of geological layers along fault surfaces perturbs both the horizontal and vertical connectivity of transmissive layers. 3D structural models that represent geological layers and structures such as faults are of prime importance for accurately representing fault displacement, trends, anisotropy of permeability and porosity heterogeneities. The rock properties greatly depend on the nature of the rocks, thus are often estimated in a coordinate sytem representing the

depositional space (Mallet, 2004; Gringarten et al., 2008).

The proposed methodology considers geological structures as a reduced set of parameters, thus opens the ways to the management of structural uncertainty in classical stochastic inversion methods like MCMC (figure 1). A simple non-conditional case study (section 5.1) shows the applicability of the proposed parameterization to reduce fault location uncertainty using a rejection sampler. Indeed, the goal is to sample the posterior distribution and obtain very different realistic models, hence rejection sampling is more appropriate than optimization techniques.

In sparse data situations, several fault arrays may be interpreted, all consistent with available data. In this context, we propose to perturb geological scenarios while maintaining consistency with interpreted data, in order to falsify scenarios that do not match two-phase flow observations (section 5.2).

## 2 Related Work and Challenges of Structural Parameters Inversion

The inverse problem is known as history matching or model calibration in the case observed data is the reservoir/aquifer dynamic flow response. The idea behind inverse modeling is obvious: if a model does not reproduce the observations, it is unlikely that its predictions are correct. Moreover, if a model matches the observations, it has a higher chance to make relevant predictions. However, only stochastic approaches provide a relevant framework for risk assessment, by considering all possible scenarios and not just one deterministic calibrated model.

Most history matching techniques focus on permeability and porosity heterogeneities, either in the field of hydrogeology (Carrera et al., 2005;

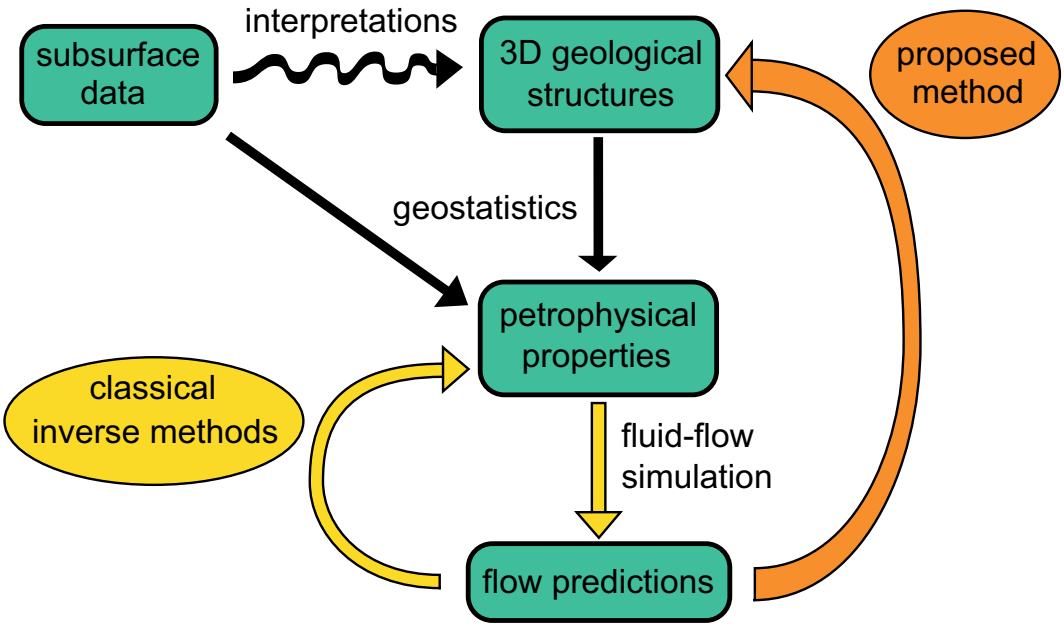


**Figure 1:** Typical modeling and inversion workflow. 3D structural modeling is based on the interpretation of subsurface data, which may entail uncertainties due to model approximations, sparse/low resolution data and possible interpretive bias. Once geological structures have been modeled, petrophysical heterogeneities are computed using geostatistical methods conditionning to subsurface data. Fluid-flow simulation then enables to compute flow predictions that can be compared to observations. The proposed parameterization opens the way to inversion strategies by updating not only petrophysical properties, as proposed by classical inverse methods, but also 3D geological structures.

de Marsily et al., 2005; Zimmerman et al., 1998), or for the hydrocarbon industry (Oliver and Chen, 2011). Other techniques focus on connectivity features (Western et al., 2001) or sedimentary objects such as channels (Ronayne et al., 2008) that may have a great impact on fluid flows by creating high-conductivity preferred flow paths. However, uncertainties about geological structures are neglected in most cases. Jenni et al. (2004) propose to handle uncertainties about fractures using gradual deformations to update field scale planar fractures while honoring a fracture density map. In this work, we propose an approach similar to Jenni et al. (2004), but

use a 3D flexible and more realistic representation of fault objects in terms of geometry and topology. Suzuki et al. (2008) focus on larger structural uncertainties. They select models that match reflection seismic data in a large set of prior structural models using a stochastic search method and the Hausdorff distance as metric to avoid running the forward problem for each ensemble member. The prior set of models is generated by perturbing the geometry of different seismic interpretations representing different geological scenarios. As Suzuki et al. (2008), we also focus on large structural uncertainties but our parameterization is flexible enough to automatically generate models with varying number of faults and fault connections.

Other uncertainties about reservoir boundaries do not only affect reservoir volume but also control the spatial layout of geological layers and thus possibly affect fluid flows. In this context, Seiler et al. (2010) use the Ensemble Kalman Filter (EnKF) to update top and bottom stratigraphic horizons of a reservoir. They use a corner-point grid to represent the most likely initial state and deform it to match simulated top and bottom horizons. As the grid is deformed, their algorithm keeps the same number of cells over realizations and thus a state vector with same dimensions as required by the EnKF.

However, none of these methods is flexible enough to sample posterior, i.e. after data inversion, realistic 3D models of geological structures. Geological surfaces (faults, fractures or stratigraphic horizons) are either planar or their geometry is limited by the gridding method, *e.g.* pillar-based grids. Our fault representation does not suffer such limitations, for implicit modeling (Calcagno et al., 2008; Frank et al., 2007; Guillen et al., 2008) is used to represent faults as isopotentials of scalar fields, namely implicit surfaces.

Inverting structural parameters using dynamic data raises both conceptual and technical issues:

1. One needs a robust method to sample structural uncertainties. While it is a common practice to generate several geostatistical realizations of facies or rock properties, few methods propose to sample uncertainties about geological structures, either geometrical (Abrahamsen, 1992; Caumon et al., 2007; Lecour et al., 2001; Mallet and Tertois, 2010; Thore et al., 2002; Wellmann et al., 2010) or topological (Cherpeau et al., 2010a; Holden et al., 2003). The algorithm used in this work is adapted from Cherpeau et al. (2010a) in order to meet the needs of inverse modeling. It generates stochastic fault networks from prior information about fault orientation, size, sinuosity and location.

2. It is not an easy task to decribe a structural model as a set of relevant parameters. In this paper, we focus on fault network parameters, considering both individual fault parameters such as orientation, size and location and also the number of faults in the model. Connections between faults are not explicitly parameterized but result from the fault network generation process (section 3).

3. From a technical standpoint, modeling realistic geological structures requires a flexible 3D representation while modeling physical phenomena requires meshes for discretization of partial differential equations describing the physical phenomenon of interest (Caumon et al., 2004). Consequently, using dynamic data to constrain geological structures calls for either back and forth and tedious conversions between the two representations, or a simplified geological representation that only enables to update the reservoir model, as proposed in the seminal work

of Holden et al. (2003). Our implicit fault surfaces can be discretized into a control volumes connectivity graph that can be fed into a flow simulator (Cherpeau et al., 2011). Hence, the geological representation does not suffer from any simplifications while the flow response can be computed from the connectivity graph.

# 3 Parameterization of Fault Networks

## 3.1 Overview of the Stochastic Fault Model

The fault generation process is a modified version of the method proposed by Cherpeau et al. (2010a,b) to sample fault-related uncertainties in sparse data situations. The algorithm sequentially simulates faults in the domain of interest. Faults are grouped into fault families representing different faulting events in terms of timing and regional stress, resulting in different input parameters about fault orientation, size, localization and sinuosity. Moreover, fault families are simulated in a given chronological order reflecting the timing of faulting events. Consequently, new simulated faults are branching faults in case of intersections with previously simulated faults. When the timing is uncertain, fault families are considered cogenetic so that truncation is randomly simulated.

Each fault object is simulated through a number of substeps and the global simulation process is as follows:

For each fault family $S$, from the oldest to the youngest:

For each fault $F$ in $S$:

1. Draw dip and strike from prior distributions.
2. Draw fault center coordinates from the fault center

probability distribution function $f$ built from input parameters (section 3.1.1).

3. Generate a fault geometry from orientation parameters, fault center and input sinuosity parameters.
4. Draw an extension along the strike direction and along the dip direction (section 3.1.2).
5. Look for intersections with previously simulated faults.

The simulation can also be conditional to interpreted fault traces. In this case, extra simulation steps assign data to faults and randomly cluster data points using geometrical concepts (Cherpeau et al., 2010b). The algorithm assigns data points to faults in different orders and generates different data combinations, hence produces different models that all honor hard data. The following sections describe some substeps of the fault object simulation.

### 3.1.1 Fault center coordinates

The fault nucleation step is a key step during the simulation process for it determines the neighborhood of a fault thus possible intersections with neighbor faults. In some geological contexts, major faults tend to repulse each other, meaning that new faults occur far from existing ones, whereas secondary faults tend to occur around major faults. If one wants to mimic such behaviors, the fault center probability distribution function $f$ may be non-uniform in the domain of interest and may depend on neighbor faults. Consequently, the fault center is drawn from $f(u, v, w)$ defined in a Cartesian grid, by successive drawing of $u \in [1; n_u]$, $v \in [1; n_v]$ and $w \in [1; n_w]$ coordinates, with $n_u, n_v, n_w$ the number of voxels along the three orthogonal axis $a_u$, $a_v$ and $a_w$ respectively. The probability $p(U = a)$ that the $u$

coordinate is equal to $a$ is:

$$\begin{aligned} p(U=a) &= \int_v \int_w f(a,v,w)dvdw \\ &= \sum_{j=1}^{n_v} \sum_{k=1}^{n_w} f(a,j,k) \end{aligned} \tag{1}$$

Once $u$ coordinate has been drawn, $p(V=b|U=a)$ with $b \in [1;n_v]$ is computed according to:

$$\begin{aligned} & p(V=b|U=a) = \int_w kf(a,b,w)dw = \sum_{k=1}^{n_w} kf(a,b,k) \\ & \text{with } k \text{ a normalization constant,} \\ & k = \frac{1}{\int_v \int_w f(a,v,w)dvdw} = \frac{1}{\sum_{j=1}^{n_v} \sum_{k=1}^{n_w} f(a,j,k)} \end{aligned} \tag{2}$$

Then, $w$ can be drawn in the same manner, *i.e.* by computing a discrete pdf along the axis $a_w$ given $u=a$ and $v=b$ coordinates. Continuous values of these discrete probability density functions are obtained by linear interpolation between discrete values, that can then be back-transformed into $x, y, z$ coordinates.

### 3.1.2 Fault extension

In order to model laterally-terminating and synsedimentary faults, the algorithm restricts the existence of a fault by a 3D ellipsoid $\mathcal{E}(x, v_1, v_2, v_3)$ with $x$ the fault center coordinates, $v_1, v_2, v_3$ three orthogonal vectors. The vectors $v_1$ and $v_2$ are respectively computed from extension values randomly drawn from length distribution along the strike direction and height distribution along the dip direction. The third vector $v_3$ is fully defined by $v_3 = v_1 \times v_2$.

This elliptical fault shape is not a restriction of our algorithm and is consistent with the litterature. Indeed, fault's tipline is considered to be circular to elliptical for an isolated fault (Barnett et al., 1987; Walsh et al., 2003) but may evolve to elongated ellipses if the vertical fault height is restricted by the thickness of a geological layer (Benedicto et al., 2003), which can be approximated by choosing $||v_1|| >> ||v_2||$.

### 3.1.3 Fault number parameter

In case of large fault-related uncertainties, not only is the fault connectivity uncertain but also the number of faults occuring in the studied area. The algorithm associates, for each fault family, a theorical number of faults $n$ with a probability of fault existence $p_e \in [0, 1]$, so that the output number of faults $m$ varies in $[0, n]$ given the following probability density function $Q(m)$:

$$Q(m) = \binom{n}{m} p_e^m \times (1 - p_e)^{n-m} \tag{3}$$

The following section presents the parameterization of our stochastic fault model to fit the needs of an inversion scheme.

## 3.2 Model Parameterization

The stochastic fault model sequentially simulates faults in the domain of interest. Each step during the simulation of a fault object stochastically set a fault parameter. Finally, each fault can be described by six parameters: dip, strike, length, height, sinuosity and center. During the fault simulation, one or more random numbers are used by the stochastic steps to generate each fault parameter, *i.e.* only one random number is used per dip, strike, height, length parameters whereas three random numbers are required to draw the fault center position. Consequently, each fault can be parameterized by a

vector with eight components $F(r_1, ..., r_8)$. A model $m$ with $n$ faults is fully parameterized by a random vector with $8n$ components $m = \{F^1, ..., F^n\}$ or $m = \{(r_1^1, ..., r_8^1); ...; (r_1^n, ..., r_8^n)\}$.

# 4 Inversion Methodology

The parameterization described in section 3.2 opens ways to inversion schemes in which geological structures are part of the inversion loop, as shown in figure 1. In this section, we present how the parameterization can be used to sample fault networks using two-phase flow response as a misfit to minimize. The inversion scheme relies on a version of the Metropolis algorithm (Metropolis et al., 1953) proposed by Mosegaard and Tarantola (1995). The Metropolis algorithm was originally designed to sample Gibbs-Boltzmann distributions in chemical science and has been since successfully applied in several fields such as mathematic, physic and biology. It is a *Markov Chain Monte Carlo* method, *Markov Chain* implicitly meaning first-order Markov chain, *i.e.* current state only depends on previous state, and *Monte Carlo* meaning that it is a random sampling method. The Metropolis algorithm is an *importance sampling* algorithm designed to solve highly nonlinear inverse problems by sampling the posterior probability density function defined by the product of the likelihood function and the prior probability density function. A general description of Monte Carlo methods as solutions to inverse problems can be found in Mosegaard and Sambridge (2002).

## 4.1 Random Walk

Inverse methods search for solutions that minimize the misfit between predictions and observations and thus need random walks or perturbation processes to explore the parameter space. In this space, the neighborhood of

point $i$ is defined as all the points $j$ that can be reached from $i$ in a single move. We define the neighborhood $\mathcal{N}_i$ of a model $m_i(r_1^i, ..., r_{8n}^i)$ in 8n-D search space as the set of all models $m_j(r_1^j, ..., r_{8n}^j)$ that can be obtained through the following steps:
For each random number $r_k^i$:

1. Compute $b_{low} = \max(0, (r_k^i(1-p)))$ with $p$ a perturbation amplitude given in input;

2. Compute $b_{up} = \min(1, (r_k^i(1+p)))$;

3. Define a uniform distribution $D_k$ $\mathcal{U}[b_{low}, b_{up}]$;

4. Randomly draw $r_k^j$ from $D_k$.

Consequently, the neighborhood $\mathcal{N}_i$ corresponds to an infinite set of models. However, any point $B$ in parameter space can be reached from any point $A$ provided a sufficient number of steps, *i.e.* the random walk covers the whole parameter space. Moreover, the proposed transition process from model $i$ to model $j$ is symmetric, *i.e.* the probability $p(m_i|m_j)$ of moving from model $m_j$ to model $m_i$ is equal to the probability $p(m_j|m_i)$ of moving from model $m_i$ to model $m_j$: $p(m_i|m_j) = p(m_j|m_i)$.

For a point $i$ in parameter space, the perturbation amplitude $p$ defines the size of the neighborhood around $i$, thus controls the speed of exploration of the parameter space. Moreover, $p$ reflects the trade-off between *exploitation* and *exploration* of the parameter space. Indeed, if $p$ is small, it means that the algorithm searches for models in a close neighborhood and thus tries to *exploit* current position in parameter space. If $p$ is large, the neighborhood is large and thus the algorithms tends to *explore* the parameter space. However, high-dimensional spaces tend to be empty spaces and the

probability of moving from a high-misfit point $i$ to a low-misfit point $j$ may be high if the move, *i.e.* the perturbation amplitude $p$, is too large.

### 4.1.1 Parameter variation speed

Another way of generating a new model could be to directly perturb the fault parameters of the current model, *i.e.* to bypass the statistical distributions. When perturbing random numbers, all numbers evolve at the same "speed" in the range $[0, 1]$ depending on the input perturbation amplitude $p$. Then, each fault parameter is drawn from the corresponding distribution, meaning that fault parameters evolve at a speed proportional to the spread of their distribution. Indeed, the narrower the distribution, the slower the parameter variation speed. Consequently, the method accounts for the prior information and uncertainty. This could be hardly achieved by perturbing directly the fault parameters.

### 4.1.2 Structural parameters evolution

The parameterization and perturbation process described above ensures that fault parameters evolve smoothly. However, the fault sinuosity is described by a single random number used as a seed for generating a correlated random noise using a geostatistical algorithm, *e.g.* Sequential Gaussian Simulation. The method is similar to Caumon et al. (2007), the generated random field is added to the scalar field representing the fault surface. The input variogram ranges needed for the geostatistical method is computed from sinuosity parameters. In contrast with gradual deformation methods, the fault sinuosity does not evolve smoothly since perturbing the seed entails the generation of a new perturbation field.

Moreover, the fault connectivity is not parameterized as it results from

the stochastic fault simulation algorithm. However, if two faults are no longer connected, the fault positions evolve smoothly so that the flow has a small volume of matrix to pass trough to connect the two disconnected faults. Consequently, even if some fault connections are lost from one model to another, the corresponding flow responses are still correlated.

#### 4.1.3 Fault number evolution

Adding or removing a fault from one model to another may entail a large perturbation of the model predictions, *e.g.* if the new or removed fault is localized in a flow-influencing zone, the flow response change may be too large to consider correlated models. In order to get smooth transitions, the existence of a fault for a given fault family is coupled with its length distribution so that the probability of a zero-length fault is $1-p_e$ with $p_e$ the probability of existence. Consequently, the length distribution is modified as illustrated in figure 2. The $\epsilon$ parameter should be chosen carefully in function of the perturbation amplitude $p$, for it controls the fault growth/contraction rate. Indeed, if $p$ is too large as compared to $\epsilon$, the fault may get its normal size or disappear during one iteration.

### 4.2 Acceptance Rule

The acceptance rule modifies the random walk (presented in section 4.1) by introducing a probability $p_{ij}$ to move from point $i$ to point $j$, so that the posterior probability density function is sampled. We use the Metropolis rule as the acceptance rule. For each model $m$, we run the forward problem, *i.e.* a flow simulation, and compute some water-cuts $w_t(m)$ at production wells. The misfit function $S(m)$ is the root mean squared error with the

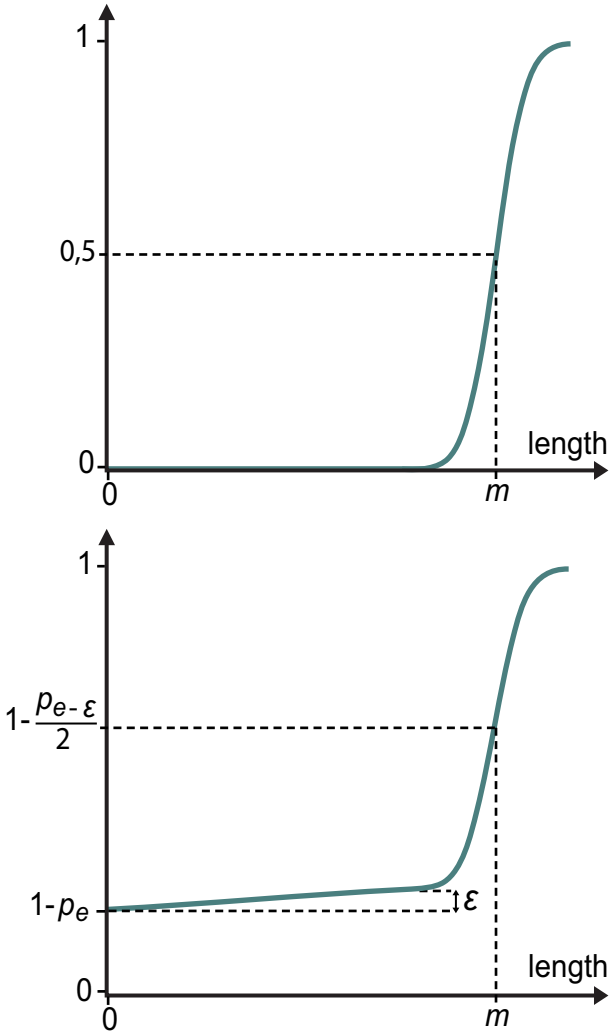


**Figure 2:** Fault length distribution coupled with existence probability parameter. At the top, an initial gaussian CDF representing the fault length with mean $m$. At the bottom, the distribution has been modified to account for the existence probability $p_e$. Parameter $\epsilon$ should be carefully set in function of the perturbation amplitude $p$.

observations $w_t^{obs}$:

$$S(m) = \sqrt{\frac{\sum_{t=1}^{l}(w_t(m) - w_t^{obs})^2}{l}} \tag{4}$$

The likelihood function is of the form $L(m) = \exp(-\frac{S(m)}{s^2})$, with $s^2$ the total noise variance. Then, the probability $p_{ij}$ of moving from point $i$ to

point $j$ is:

$$\begin{aligned} p_{ij} &= 1 && \text{if } L(m_j) \geq L(m_i) \\ &= \frac{L(m_{j)}}{L(m_i)} && \text{otherwise} \\ &= \exp(-\frac{S(m_j) - S(m_i)}{s^2}) \end{aligned} \tag{5}$$

Note that the misfit function could be any function of interest for the study at hand. We here use it for the two examples presented in section 5. Moreover, the methodology could be used in other scientific areas. It could for instance be used in geophysical inversions to better constrain subsurface structure. In this case, a wave propagation method would replace the flow simulation during the forward modeling step in the inverse methodology.

# 5 Applications

## 5.1 Updating Prior Structural Uncertainty

This case study aims at understanding the behavior of the method for a simple case and assessing the amount of structural information introduced by a simple history matching problem. Fault occurence and location are not constrained by any interpreted data.

### 5.1.1 Input Data

The studied area of size 1400x1400x200m has one injector and one producer wells on both sides of a fault zone composed of two fault families:

1. The first fault family corresponds to major faults aligned along the main flow path:

    - strike: uniform, $[110 - 130]^\circ$
    - dip: Gaussian, $mean = 70^\circ$, $\sigma = 3^\circ$

- length: Gaussian, $mean = 800m$, $\sigma = 50m$
- height: Gaussian, $mean = 350m$, $\sigma = 5m$
- number of faults: 3
- probability of existence $p_e = 0.8$
- sinuosity wavelength along strike direction: 500m
- sinuosity wavelength along dip direction: 500m

2. The second fault family corresponds to branching faults:

   - strike: uniform, $[50 - 70]^\circ$
   - dip: Gaussian, $mean = 70^\circ$, $\sigma = 3^\circ$
   - length: Gaussian, $mean = 500m$, $\sigma = 30m$
   - height: Gaussian, $mean = 300m$, $\sigma = 5m$
   - number of faults: 2
   - probability of existence $p_e = 0.7$
   - sinuosity wavelength along strike direction: 350m
   - sinuosity wavelength along dip direction: 300m

The total number of faults varies in [0,5] according to equation 3. The matrix has a constant porosity of 0.2 and a constant permeability of 100mD. Faults are considered flow barriers with a permeability set at 0.01mD, hence uncertainty about fault transmissibility is not investigated.

### 5.1.2 Inversion Scheme

The Metropolis algorithm is designed to sample asymptotically the posterior distribution function so that the starting point or seed to generate the Markov chain of models is meaningless (Mosegaard and Tarantola, 1995).

Then, after the burn-in time is over, samples of the posterior can be picked according to a "waiting time" to get independent samples.

To speed up the process, we propose to generate first 2000 samples from the prior and then to use the 100 models that best match flow data as seeds for generating 100 independent Markov chains of models using the acceptance rule described in section 4.2. Each chain stops when the misfit is below a given threshold or 500 models are generated. This process does not guarentee unbiased sampling of the posterior distribution, but corresponds to a practical use of the method in which only models deemed acceptable are retained for further use.

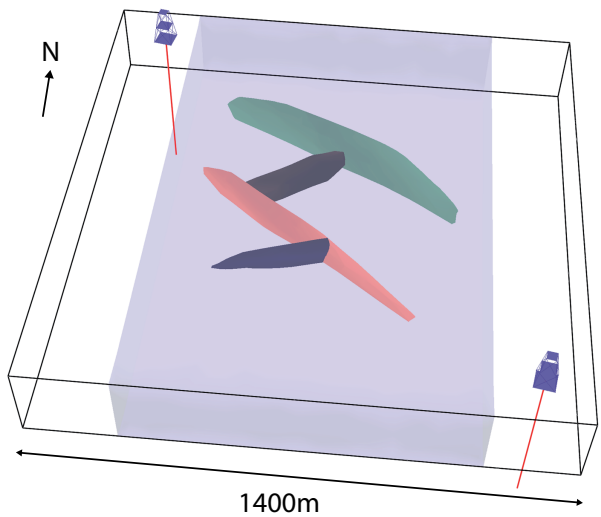


**Figure 3:** The reference model has 4 faults. The shaded zone represents the fault zone thus ensures faults to be inside this zone.

### 5.1.3 Reference Model and selected models

The reference model is randomly drawn among the 2000 prior models and is presented in figure 3. Then, the misfit is computed for other prior models using equation 4 and water-cut data. Results are presented in figure 4.

The misfit threshold to accept posterior models is set to 10 (from 51 for the worst of the 100 selected models); only four models are below that threshold among the 2000 models.

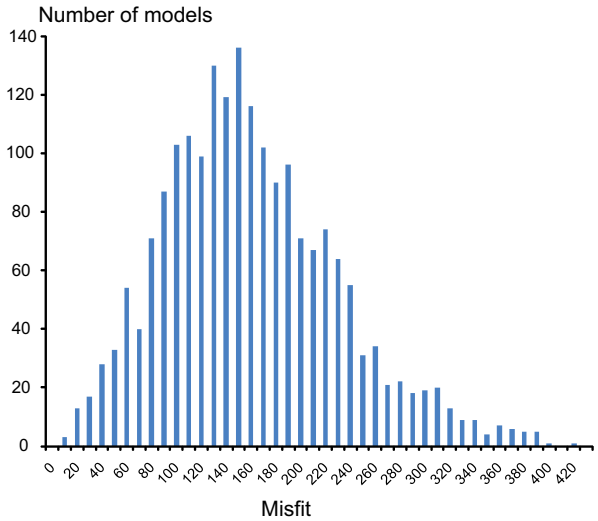


**Figure 4:** Mifsit repartition computed from equation 4. Only the 100 selected models are used as seeds to generate Markov chain of models and four of them have a misfit below 10 thus are considered as posterior models.

### 5.1.4 Sensitivity to Perturbation Amplitude Parameter

The perturbation amplitude parameter $p$ controls the exploitation versus exploration of the parameter space. To test the sensitivity of our methodology to parameter $p$, the best model is perturbed using different values of $p$. 100 models are generated for each value of $p$, and their flow-response misfit is computed and summarized in figure 5. Statistics are almost constant for small perturbations ranging from 0.0001 to 0.03 maybe due to the impact of the fault sinuosity seed that generates a new sinuosity whatever the perturbation value. Moreover, maximum values are unexpectedly constant, maybe due to a fault configuration in unstable state, *e.g.* a fault center coordinate

close to a neighbor fault may switch of side even for small changes. Minimum misfit values are also almost constant due to the perturbation process that does not exclude small perturbations even for large $p$ values. Consequently, we focus on centile values that do not depend on extrem values. For $p$ ranging from 0.04 to 1, the median misfit increases linearly with a correlation coefficient of 0.997.

These results greatly depend on prior information, thus cannot be generalized. Indeed, as described in section 4.1.1, the parameters evolution speed depends on the parameters distribution, *e.g.* a Dirac distribution always generates the same value thus the speed is null whatever the perturbation amplitude is. Given these results, the perturbation amplitude $p$ is set to 0.03 for the case study which ensures small misfit changes while enabling exploration of the parameter space.

#### 5.1.5 Results

Any perturbation among the fault parameters may entail a fault network topology change and impact the flow response. Moreover, parameters are connected, *e.g.* a change of fault position increasing the distance between two faults may be balanced by a fault growth decreasing the distance between the two faults, thus cannot be considered individually. In case the fault size is small as compared to other faults, changes of fault location or orientation may entail unsignificant changes in flow response. Consequently, we propose to compare fault occurence probability maps (figures 6 and 7) instead of posterior distributions.

Only three Markov chains among the 100 chains did not manage to reduce the misfit under the input threshold. An example of misfit evolution for a seed model is shown in figure 8. The average acceptance rate is about 43%

for all Markov chains, which is relevant with prior studies (Gelman et al., 1996; Mosegaard and Sambridge, 2002). Both prior probability maps do not show any particular feature but a large zone with medium probability. Posterior probability maps show specific features that are present in the two subsets which suggests relative convergence and stability of results.

Posterior probability maps relatively well resolve fault location uncertainty. A typical realization would have two faults of family $A$ on both sides of the main flow path and one fault of family $B$ playing the role of flow barrier along the flow path, which is relatively similar to the reference model. Other faults located elsewhere in the model do not impact the flow response, which may explain why these areas are not very well resolved on the posterior probability maps. Indeed, any hard data constrains fault location and the impact of conditioning is still to be evaluated.

## 5.2 Analysis of Structural Scenarios

The interpretation of subsurface data may lead to different interpretations (Bond et al., 2007) and thus different geological scenarios. Fluid-flow data represent additional data that may help to analyze different interpretations and to falsify those that do not match observations (Tarantola, 2006). In this section, we propose to analyze different structural scenarios built from the interpretation of seismic sections by generating a family of models around each interpretation. The sinuosity and size of each deterministic fault of the initial interpretation are perturbed. Other fault parameters such as fault center location and orientation are frozen. The initial interpretations are used as seeds to generate Markov chains of models as in section 5.1 and conditioning to seismic interpretations is made during neighborood search. Only models that match flow data are retained and other interpretations

are falsified. The study is not meant to be statistical but practical, hence scenarios that do not have any neighbor model with an acceptable misfit to water-cut data after a large number of iterations are falsified.

### 5.2.1 Data Overview

Eight structural scenarios have been built from four interpreted cross-sections, one of them is considered as the reference scenario (figure 9). There are one injector in the center and four producers in the corners of the studied area of size 1400x1400x200m. As in section 5.1, the misfit is the root mean squared error (equation 4) of water-cut data. Matrix porosity is set at 0.2, matrix permeablity is 5mD and fault permeability is 0.01mD.

### 5.2.2 Results

Initial interpretations are perturbed until an acceptable misfit is reached. This arbritary choice is motivated by a practical need to obtain a match in limited time, a more rigorous version of MCMC could be run in principle to better sample the search space, at the expense of computational efficiency.

The evolution of the misfit from one model to another is presented in figure 10. The average acceptance rate for all scenarios is about 33%. Scenarios $I_1, I_4, I_5$ and their corresponding chain of models have a flow-response that remains far from the reference thus are falsified. Scenarios $I_0, I_2, I_3, I_6$ have a neighbor model within an acceptable misfit range, respectively $I'_0, I'_2, I'_3, I'_6$ (figure 11), which are considered consistent with available water-cut data.

Conclusions would have been much different by only considering the initial interpretations misfit, *e.g.* interpretations $I_2, I_3, I_6$ have a larger initial misfit than falsified interpretations $I_1, I_4, I_5$.

The low-misfit models $I'_0, I'_2, I'_3, I'_6$ can then be used for further modeling

steps with a higher confidence into their corresponding structural models.

## 6 Discussion and Conclusions

The proposed parameterization of faults enables to sample posterior fault networks with various number of faults and fault connections. The modeling method uses a realistic 3D representation of fault surfaces that goes beyond planar elliptical surfaces and enables to model realistic fault arrays. It considers faults as a random vector corresponding to fault parameters and opens the way to inversion of geological structures using realistic descriptions. It goes beyond traditional inversion schemes by pertubing both fault geometry and fault network topology.

The first presented example shows how fault position uncertainty can be reduced using our parameterization and a version of the Metropolis algorithm to sample posterior models. The sampled 100 posterior models suggest specific features about fault location. The second example shows the interest of the methodology when conditioning data is available, by analyzing structural scenarios and falsifying those that do not have any neighbor model consistent with observations. The simulation method enables to permanently honor data points during the perturbation process, which reduces the neighborhood around selected scenarios and ensures the consistency of the proposed models.

The smooth evolution of fault parameters, even for the number of faults coupled with size evolution, enables to generate correlated flow responses from one sample to another. Only the fault sinuosity does not evolve smoothly from one sample to another but gradual deformation could help to obtain more continuous transitions (Hu, 2000). Other fault-related uncertainties could be considered to complete the method, such as fault trans-

missibility which may have a significant impact on the flow response. In the presented examples, faults occur in a homogeneous geological layer but rock heterogenetities could also be taken into account. In such cases, fault displacement should also be modeled for it may connect or disconnect transmissive layers. This point will be addressed in a future contribution.

In our first simple MCMC example, the forward problem, *i.e.* the flow simulation, represents about 91% of the computation time (about 9% for fault simulation, the time for discretization and generation of flow input files is negligible) for an average of 1min20s per MCMC iteration on a 3.4Ghz PC with about 8000 flow elements and 16000 connections per discretized model. When complexity increases as in the second example with a larger number of faults and wells, this computation time is even more important (about four times longer than first example) even if time proportions remain relatively constant: the inversion spends 88% of the computation time to solve flow equations, 12% to simulate faults (maybe due to extra simulation steps needed for data conditioning, other steps are insignificant). These performance issues and the large number of parameters do not enable to rigorously sample the posterior. The presented inversion strategy is parallel by nature since Markov chains of models can be computed independently, which reduces the total computation time of the presented examples. Other methodologies could also be used in order to reduce the computation time, *e.g.* Mariethoz et al. (2010) propose a stochastic stopping criterion to reduce the number of forward runs. Distance-based optimization (Sambridge, 1999; Scheidt and Caers, 2009; Suzuki et al., 2008) could also be used either separately or in conjunction with the proposed method for more efficient sampling.

Interpretations of inversion results are difficult to analyze due to the

amount of data and correlation between parameters. As argued by Mosegaard and Tarantola (1995), there is still a need for methods to summarize information and reveal specific features of posterior models, that go beyond the computation of posterior distributions of individual inverse parameters. In that sense, Wellmann and Regenauer-Lieb (in press) propose to use information entropy to compare results and evaluate uncertainty reduction of geological units repartition after data inversion.

Beyond practical solutions and issues about the computation cost of inversions, one may wonder whether the search for fully sampled posterior distributions is meaningful. In the case of structural modeling, one only needs some models roughly covering the uncertainty space to continue the modeling workflow and make predictions. The Metropolis sampler enables such spread over the posterior probability function that cannot be obtained with optimization techniques. If other uncertainties are to be taken into account, *e.g.* about fault rock properties or fault slip, the empty space problem becomes even more difficult to handle and calls for much more models, which makes the computation time quickly becoming unreasonable in practice.

## Acknowledgments

The authors would like to acknowledge the Gocad Consortium for funding this research as well as Stanford University for providing the GPRS flow simulator. This work is partially supported by the European Research Council (ERC-StG-205693).

## References


P. Abrahamsen. Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In A. Soares, editor, *Geostatistics Tria '92. Volume 1. Forth International Geostatistical Congress. Tria, Portugal, 13-18 September, 1992*, volume 5 of *Qualitative Geology and Geostatistics*, pages 385–398. Kluwer Academic Publ., 1992.

J. A. M. Barnett, J. Mortimer, J. H. Rippon, J. J. Walsh, and J. Watterson. Displacement geometry in the volume containing a single normal fault. *AAPG Bulletin*, 71(8):925–937, 1987.

A. Benedicto, R. A. Schultz, and R. Soliva. Layer thickness and the shape of faults. *Geophys. Res. Lett.*, 30(20), 2003. doi: 10.1029/2003GL018237.

C. Bond, A. Gibbs, Z. Shipton, and S. Jones. What do you think this is? "conceptual uncertainty" in geoscience interpretation. *GSA Today*, 17 (11):4–10, 2007.

P. Calcagno, J. Chilès, G. Courrioux, and A. Guillen. Geological modelling from field data and geological knowledge: Part i. modelling method coupling 3d potential-field interpolation and geological rules. *Physics of the Earth and Planetary Interiors*, 171(1-4):147 – 157, 2008. ISSN 0031-9201. Recent Advances in Computational Geodynamics: Theory, Numerics and Applications.

J. Carrera, A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten. Inverse problem in hydrogeology. *Hydrogeology Journal*, 13:206–222, 2005. ISSN 1431-2174. 10.1007/s10040-004-0404-7.

G. Caumon, O. Grosse, and J.-L. Mallet. High resolution geostatistics on coarse unstructured flow grids. In O. Leuangthong and C. V. Deutsch, ed-

itors, *Geostatistics Banff, Proc. of the seventh International Geostatistics Congress*. Kluwer, Dordrecht, 2004.

G. Caumon, A.-L. Tertois, and L. Zhang. Elements for stochastic structural perturbation of stratigraphic models. In *Proc. Petroleum Geostatistics*. EAGE, 2007.

N. Cherpeau, G. Caumon, and B. Lévy. Stochastic simulations of fault networks in 3D structural modeling. *Comptes Rendus Geoscience*, 342 (9):687 – 694, 2010a. ISSN 1631-0713. doi: 10.1016/j.crte.2010.04.008.

N. Cherpeau, G. Caumon, and B. Lévy. Stochastic simulation of fault networks from 2D seismic lines. *SEG Technical Program Expanded Abstracts*, 29(1):2366–2370, 2010b. doi: 10.1190/1.3513325.

N. Cherpeau, G. Caumon, J. Caers, and B. Lévy. Assessing the impact of fault connectivity uncertainty in reservoir studies using explicit discretization. In *SPE Reservoir Characterization and Simulation Conference and Exhibition (RCSC)*, 2011. SPE 148085.

G. de Marsily, F. Delay, J. Gonalvs, P. Renard, V. Teles, and S. Violette. Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13:161–183, 2005. ISSN 1431-2174. 10.1007/s10040-004-0432-3.

T. Frank, A.-L. Tertois, and J.-L. Mallet. 3D-reconstruction of complex geological interfaces from irregularly distributed and noisy point data. *Computers & Geosciences*, 33(7):932 – 943, 2007. ISSN 0098-3004.

A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York, 1996.

E. Gringarten, B. Arpat, S. Jayr, and J.-L. Mallet. New Geologic Grids for Robust Geostatistical Modeling of Hydrocarbon Reservoir. In J. Ortiz and X. Emery, editors, *Proc. eighth Geostatistical Geostatistics Congress*, volume 2, pages 647–656. Gecamin ltd, 2008.

A. Guillen, P. Calcagno, G. Courrioux, A. Joly, and P. Ledru. Geological modelling from field data and geological knowledge: Part ii. modelling validation using gravity and magnetic data inversion. *Physics of the Earth and Planetary Interiors*, 171(1-4):158 – 169, 2008. ISSN 0031-9201. Recent Advances in Computational Geodynamics: Theory, Numerics and Applications.

L. Holden, P. Mostad, B. F. Nielsen, J. Gjerde, C. Townsend, and S. Ottesen. Stochastic structural modeling. *Math. Geol.*, 35(8):899–914, 2003.

L. Hu. Gradual deformation and iterative calibration of gaussian-related stochastic models. *Mathematical Geology*, 32:87–108, 2000. ISSN 0882-8121. doi: 10.1023/A:1007506918588.

S. Jenni, L. Y. Hu, R. Basquet, G. de Marsily, and B. Bourbiaux. History matching of stochastic models of field-scale fractures: Methodology and case study. In *SPE Annual Technical Conference and Exhibition*, Houston, Texas, 2004.

M. Lecour, R. Cognot, I. Duvinage, P. Thore, and J.-C. Dulac. Modeling of stochastic faults and fault networks in a structural uncertainty study. *Petroleum Geoscience*, 7:S31–S42, 2001.

J.-L. Mallet. Space-time mathematical framework for sedimentary geology. *Mathematical geology*, 36(1):1–32, 2004.

J.-L. Mallet and A.-L. Tertois. Solid earth modeling and geometric uncertainties. In *SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.

Manzocchi, T., Walsh, J. J., Nell, P. A. R., and Yielding, G. Fault transmissibility multipliers for flow simulation models. *Pet. Geol.*, 5:53–63, 1999.

G. Mariethoz, P. Renard, and J. Caers. Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research*, 46, 2010. doi: 10.1029/2010WR009274.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

K. Mosegaard and M. Sambridge. Monte Carlo analysis of inverse problems. *Inverse Problems*, 18(3):R29–R54, 2002.

K. Mosegaard and A. Tarantola. Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7):12431–12447, 1995. doi: 10.1029/94JB03097.

D. Oliver and Y. Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15:185–221, 2011. ISSN 1420-0597. 10.1007/s10596-010-9194-2.

M. J. Ronayne, S. M. Gorelick, and J. Caers. Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach. *Water Resources Research*, 44, W08426, 2008. doi: 10.1029/2007WR006635.

M. Sambridge. Geophysical inversion with a neighbourhood algorithm-I: Searching a parameter space. *Geophysical Journal International*, 138: 479–494, 1999.

C. Scheidt and J. Caers. Representing spatial uncertainty using distances andkernels. *Mathematical Geosciences*, 41:397–419, 2009. ISSN 1874-8961. doi: 10.1007/s11004-008-9186-0.

A. Seiler, S. Aanonsen, G. Evensen, and J. C. Rivens. Structural Surface Uncertainty Modeling and Updating Using the Ensemble Kalman Filter. *SPE Journal*, 15(4):1062–1076, 2010. ISSN 1086-055X.

S. Suzuki, G. Caumon, and J. Caers. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational Geosciences*, 12:105–119, 2008.

A. Tarantola. Popper, bayes and the inverse problem. *Nature Physics*, 2: 492–494, 2006.

P. Thore, A. Shtuka, M. Lecour, T. Ait-Ettajer, and R. Cognot. Structural uncertainties: determination, management and applications. *Geophysics*, 67(3):840–852, 2002.

J. J. Walsh, W. R. Bailey, C. Childs, A. Nicol, and C. G. Bonson. Formation of segmented normal faults: a 3-D perspective. *Journal of Structural Geology*, 25(8):1251 – 1262, 2003. ISSN 0191-8141. doi: 10.1016/S0191-8141(02)00161-X.

J. F. Wellmann and K. Regenauer-Lieb. Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. *Tectonophysics*, in press. ISSN 0040-1951. doi: 10.1016/j.tecto.2011.05.001.

J. F. Wellmann, F. G. Horowitz, E. Schill, and K. Regenauer-Lieb. Towards incorporating uncertainty of structural data in 3d geological inversion. *Tectonophysics*, 490(3-4):141 – 151, 2010. ISSN 0040-1951. doi: 10.1016/j.tecto.2010.04.022.

A. W. Western, G. Blöschl, and R. B. Grayson. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, 37(1):83+, 2001. ISSN 0043-1397. doi: 10.1029/2000WR900241.

D. A. Zimmerman, G. de Marsily, C. A. Gotway, M. G. Marietta, C. L. Axness, R. L. Bras, J. Carrera, P. B. Davies, D. P. Gallegos, J. G. Hernández, A. L. Gutjahr, P. K. Kitanidis, A. M. Lavenue, D. McLaughlin, B. S. RamaRao, C. Ravenne, R. L. Beauheim, Y. Rubin, S. P. Neuman, P. Grindrod, and G. Dagan. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resources Research*, 34(6): 1373–1413, 1998.

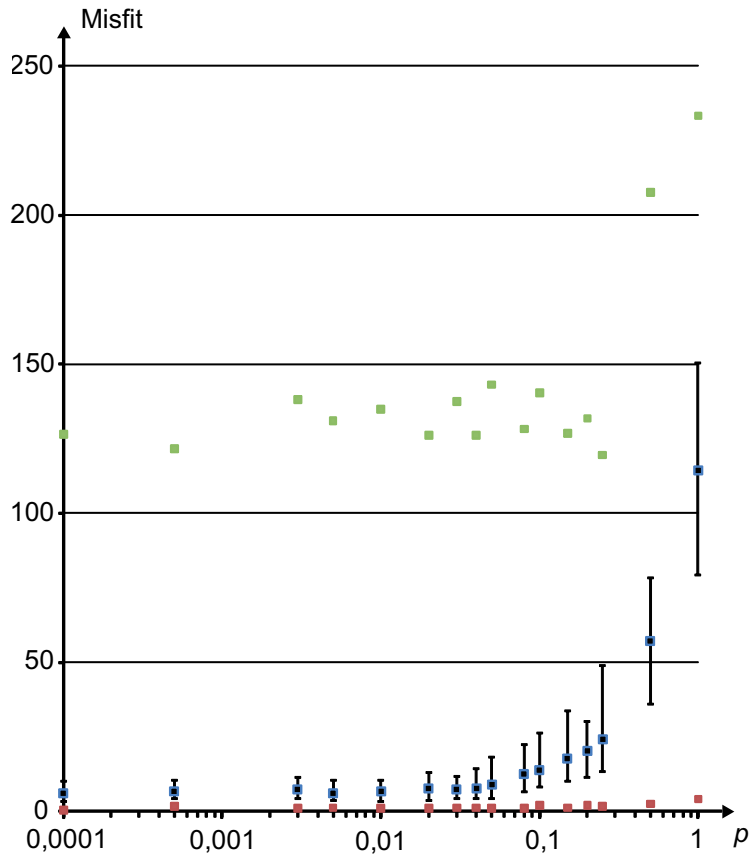


**Figure 5:** Misfit values in function of perturbation amplitude $p$ in logarithmic scale. Vertical bars represent interquartile ranges with median value. For $p$ values ranging from 0.0001 to 0.03, median values are almost constant, then increase with larger interquartile ranges. Minimum (red squares) and maximum (green squares) values are almost constant for all $p$ values due to the perturbation process and maybe unstable fault position.

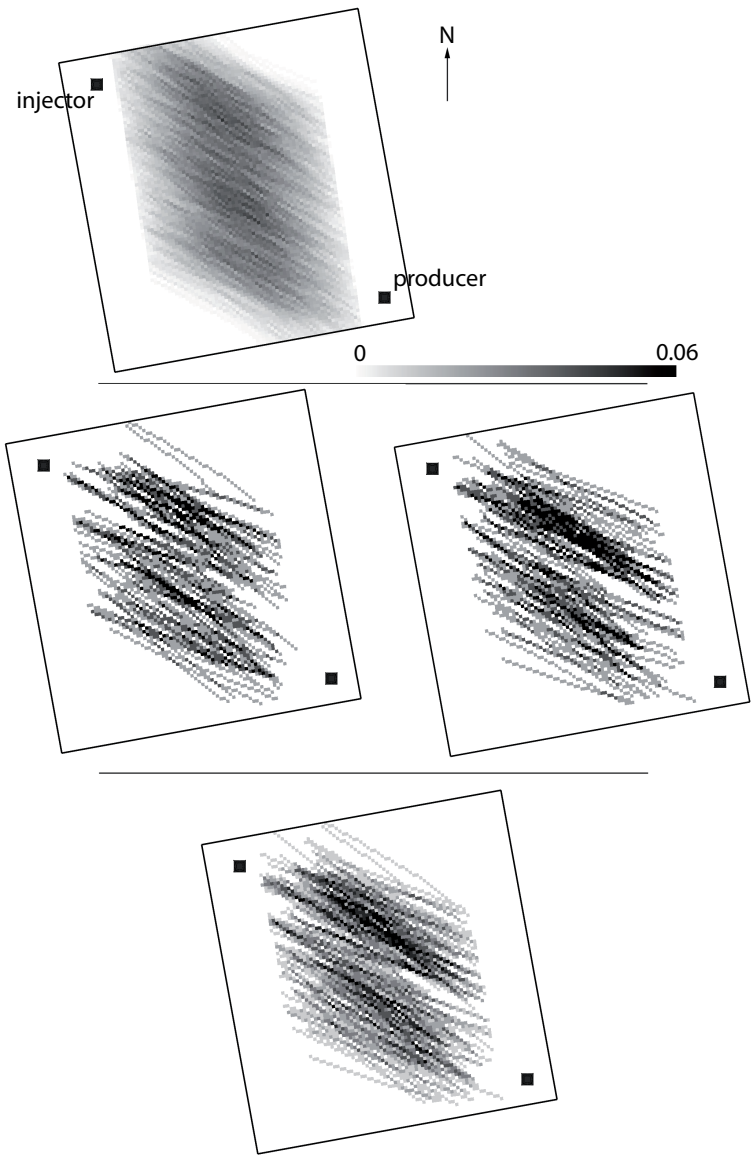


**Figure 6:** Probability maps of faults belonging to family $A$. Top: the map corresponds to the 2000 prior models. Middle: posterior models computed using the inversion scheme from the best 50 models (left) and from the 50 following models (right). The two sets of models show similar characteristics, *i.e.* two high-probability areas on both sides of the main flow path. Bottom, the probability map computed from the 100 posterior models.

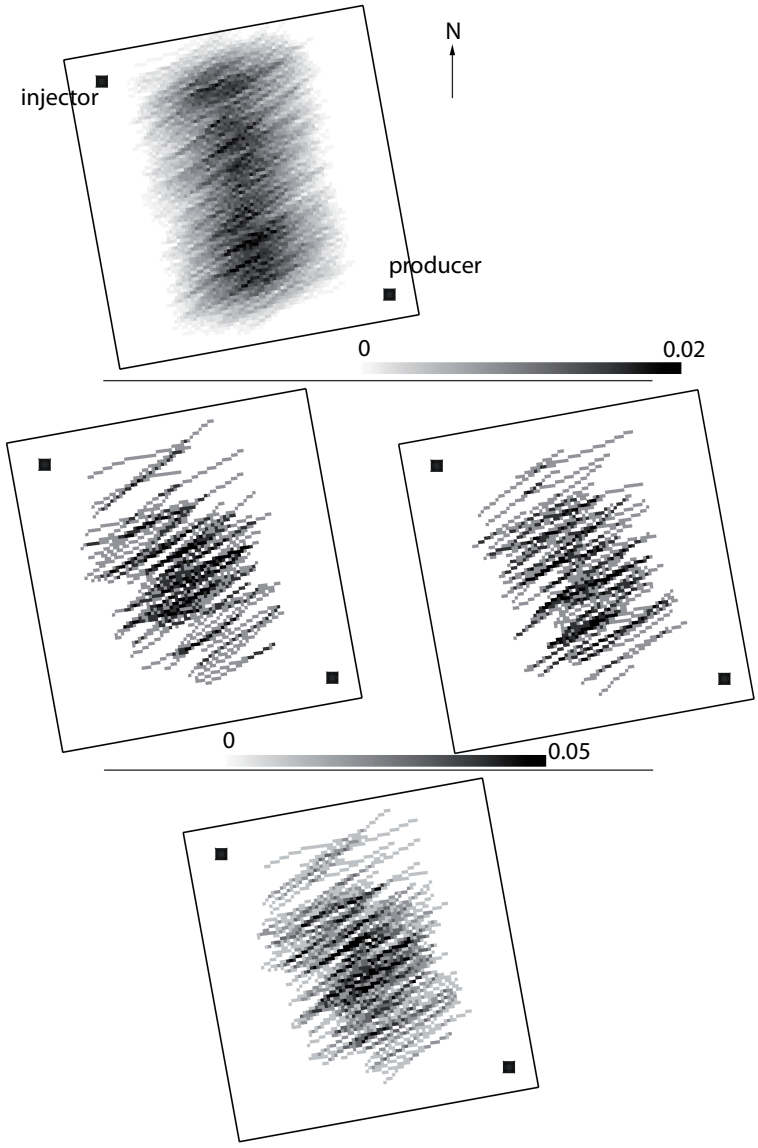


**Figure 7:** Probability maps of faults belonging to family $B$. Top: the map obtained from the 2000 prior models. Middle: posterior models computed using the inversion scheme from the best 50 models (left) and from the 50 following models (right). Bottom: posterior probability map built from the 100 posterior models. The posterior map highlights an area on the main flow path which is related to the reference with a fault playing the role of flow barrier.

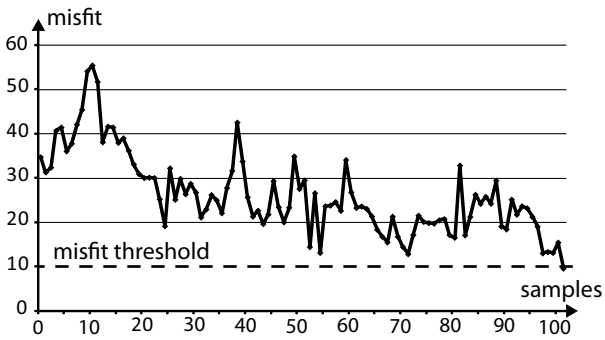


**Figure 8:** Misfit evolution for one seed model. The Markov chain stops when the misfit is below the input threshold and the last sample is used to compute posterior statistics.

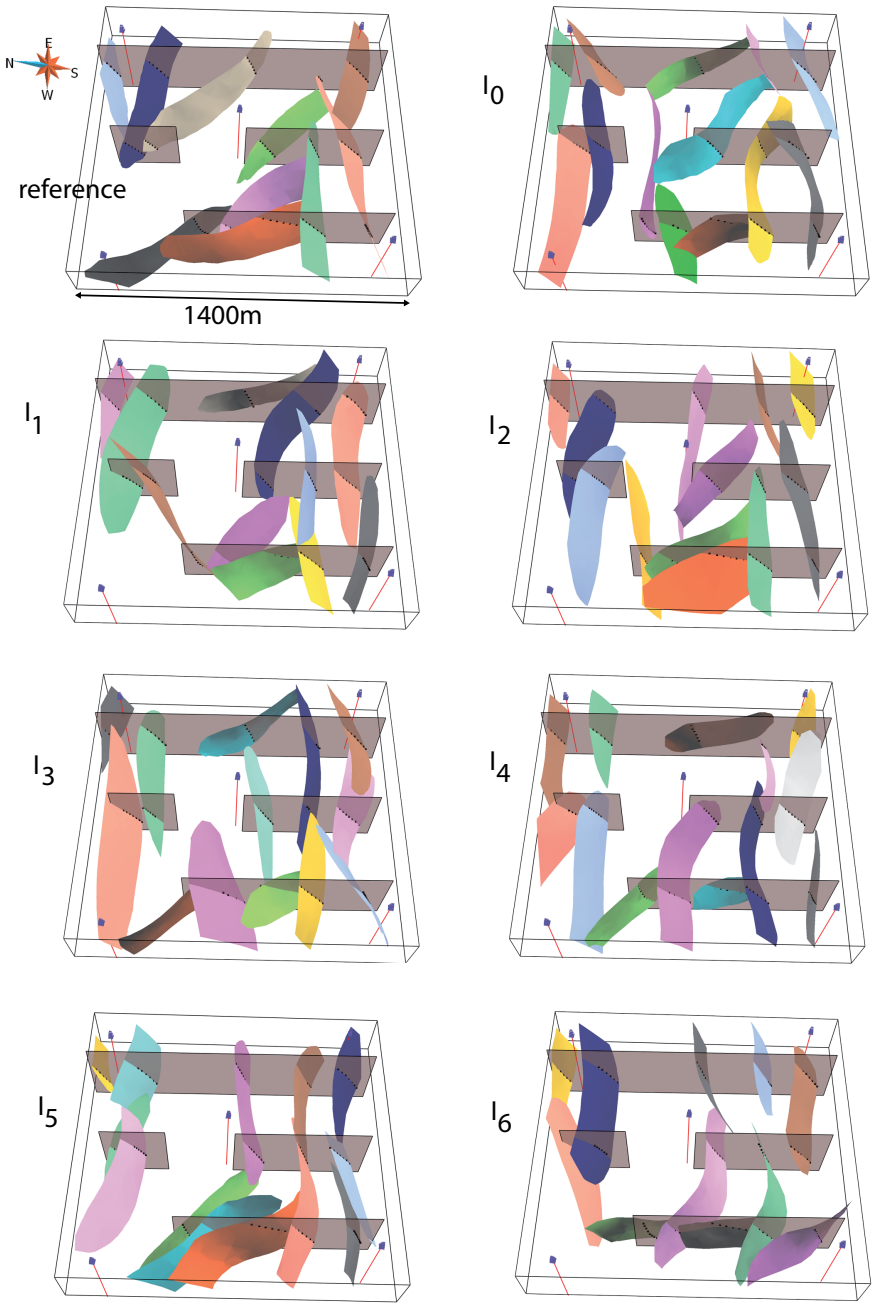


**Figure 9:** Initial interpretations and reference. Only fault sinuosity and fault size are perturbed during the inversion scheme, other parameters are frozen.

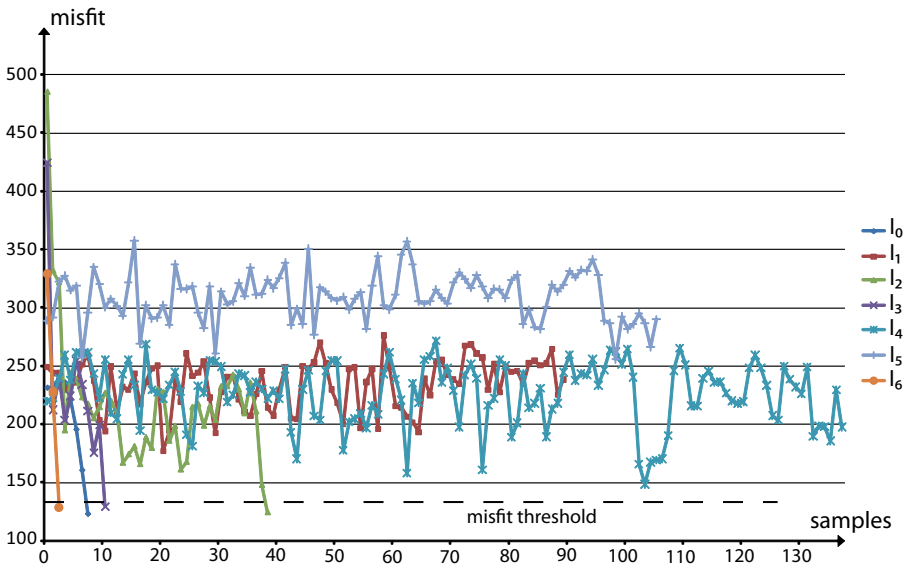


**Figure 10:** Misfit evolution of seven structural scenarios using a version of Metropolis algorithm proposed by Mosegaard and Tarantola (1995) and water-cut data. Scenarios $I_0, I_2, I_3, I_6$ have a neighbor model consistent with available data. Other scenarios $I_1, I_4, I_5$ are falsified.

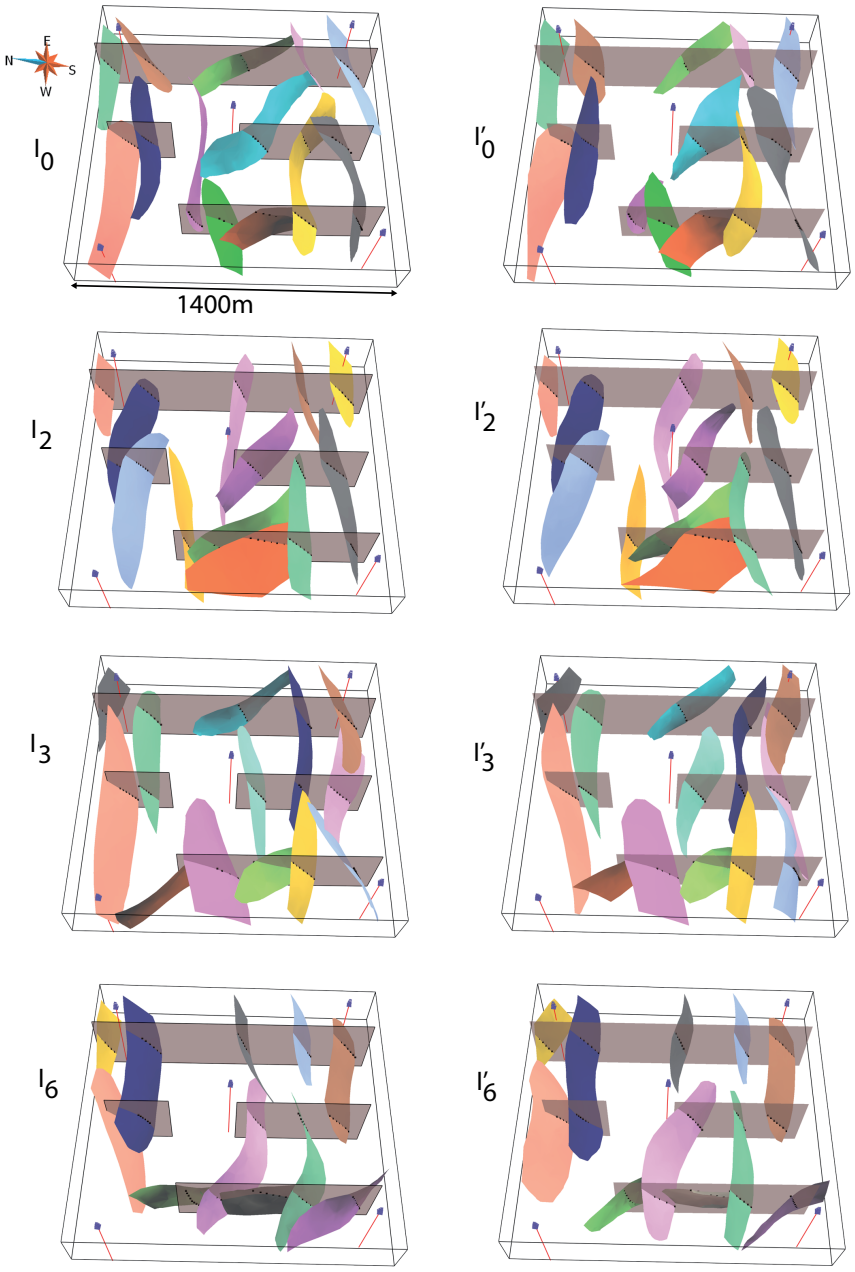


**Figure 11:** Left: initial interpretations used as seeds for generating Markov chains of models. Right: the end-member of each chain, consistent with water-cut data. Other scenarios did not produce any acceptable model thus have been falsified.