Author's version, *JGR: Solid Earth* 126, e2020JB020022, https://doi.org/10.1029/2020JB020022

Multi-scenario interpretations from sparse fault evidence using graph theory and geological rules

Gabriel Godefroy [1], Guillaume Caumon [1], Gautier Laurent [1,2], and François Bonneau [1]

[1] Université de Lorraine, CNRS, GeoRessources, ENSG
[2] Univ. Orléans, CNRS, BRGM, ISTO, UMR 7327
[1]F-54000 Nancy, France
[2]F-45071, Orléans, France

Key Points:

- Several plausible scenarios can be made when interpreting faulted structures from sparse subsurface data.
- From numerical rules expressing conceptual knowledge, a graph-based sampler generates several possible fault scenarios honoring spatial data.
- Numerical experiments suggest that the use of coherent interpretation rules increases the likelihood of generating correct interpretations.

Corresponding author: Gabriel Godefroy, Gabriel.Godefroy@hotmail.fr

Abstract

The characterization of geological faults from geological and geophysical data is often subject to uncertainties, owing to data ambiguity and incomplete spatial coverage. We propose a stochastic sampling algorithm which generates fault network scenarios compatible with sparse fault evidence while honoring some geological concepts. This process is useful for reducing interpretation bias, formalizing interpretation concepts, and assessing first-order structural uncertainties. Each scenario is represented by an undirected association graph, where a fault corresponds to an isolated clique, which associates pieces of fault evidence represented as graph nodes. The simulation algorithm samples this association graph from the set of edges linking the pieces of fault evidence that may be interpreted as part of the same fault. Each edge carries a likelihood that the endpoints belong to the same fault surface, expressing some general and regional geological interpretation concepts. The algorithm is illustrated on several incomplete data sets made of three to six two-dimensional seismic lines extracted from a three-dimensional seismic image located in the Santos Basin, offshore Brazil. In all cases, the simulation method generates a large number of plausible fault networks, even when using restrictive interpretation rules. The case study experimentally confirms that retrieving the reference association is difficult due to the problem combinatorics. Restrictive and consistent rules increase the likelihood to recover the reference interpretation and reduce the diversity of the obtained realizations. We discuss how the proposed method fits in the quest to rigorously (1) address epistemic uncertainty during structural studies and (2) quantify subsurface uncertainty while preserving structural consistency.

Plain Language Summary

This paper presents a way to generate interpretation scenarios for geological faults from incomplete spatial observations. The method essentially solves a "connect the dots" exercise that honors the observations and geological interpretation concepts formulated as mathematical rules. The goal is to help interpreters to characterize how the lack of data affects geological structural uncertainty. The proposed method is original in the sense that it does not anchor the scenarios on a particular base case, but rather uses a global characterization formulated with graph theory to generate possible fault network interpretations. The application on a faulted formation offshore Brazil where observations have been decimated, shows that the method is able to consistently generate a set of interpretations encompassing the interpretation made from the full data set. It also highlights the computational challenge of the problem and the difficulty to check the results in settings where only incomplete observations exist. The proposed method, however, opens novel perspectives to address these challenges.

Introduction

In structural characterization of geophysical data and geological mapping, the lack of conclusive observations generally makes interpretation necessary to obtain a consistent subsurface model. Indeed, geological observations and geophysical signals often have an incomplete spatial coverage and non-unique interpretations due to a lack of resolution or physical ambiguities (Wellmann & Caumon, 2018). As a result, structural uncertainty often affects fundamental research on earth's structure such as, for example, the understanding of rift development and earthquake processes (Mai et al., 2017; Riesner et al., 2017; Sepúlveda et al., 2017; Zakian et al., 2017; Gombert et al., 2018; Ragon et al., 2018; Tal et al., 2018). Faults are important elements of such studies because they have distinct hydromechanical properties and because the accumulation of fault slip over geological time directly impacts the geometrical layout of rock. Understanding fault uncertainty is, therefore, essential in many geoscience studies, but faces challenges as human-based interpretations generally aim at producing one or a few interpretation(s) deemed most likely. Computer-based models have a potential, at the expense of some simplifications, to explore a larger set of acceptable scenarios which can then be scrutinized by experts or used as prior model space for inverse methods (Tarantola, 2006; de la Varga & Wellmann, 2016). This paper describes such a computer-based sampling method to interpret fault structures from sparse data.

Faults are typically inferred from observations made on outcrops, wells, geophysical images, or through the inversion of focal mechanisms. Classically, the interpretation of these data is translated into points, lines or surfaces indicating the fault position and orientation. The problem of understanding fault structures from such incomplete geometric interpretations (*fault evidence* or *fault data*) is a classical structural geology exercise. Mapping or modeling geological faults from sparse observations can be divided into four main steps: (0) The choice of an interpretation concept, often based on the tectonic setting; (1) The *fault data association problem* (also termed fault correlation by Freeman et al., 1990), which aims at determining which of the pieces of evidence may belong to the same fault (Figure 1a,b,c); (2) The interpolation problem, which determines fault geometry and displacement from available data, and which has been extensively addressed in deterministic geological modeling (see Section 3 of Wellmann & Caumon, 2018, and references therein); (3) Optionally, the simulation of unobserved structures can be addressed by appropriate statistical point processes (Aydin & Caers, 2017; Cherpeau et al., 2010b; Holden et al., 2003). Even though items (2) and (3) present many unresolved problems, this contribution focuses on the data association problem, which has received less attention and is a prerequisite to address the other challenges.

This paper proposes a computational method that accounts for uncertainty while solving the data association problem arising while interpreting sparse subsurface data such as 2D cross-sections. We build on a recent formalism (Godefroy et al., 2019), where an association scenario is represented by an undirected graph ($\mathbb{G}^{\mathbf{asso}}$, Figure 1.e). In this graph, each node represents a piece of fault evidence. The nodes can carry information regarding the observation in the form of label (e.g., normal or reverse fault) or weights (e.g., fault throw value). Edges associates nodes belonging to the same fault, as illustrated in Figure 1.

A key idea of Godefroy et al. (2019) is to define $\mathbb{G}^{\mathbf{asso}}$ as a subset of a much larger list of associations $\mathbb{G}^{\mathbf{all}}$ containing all the potential pairwise associations for the available pieces of evidence, for each family (Figure 1.d). These potential edges carry weights representing the likelihood that any pair of fault data belongs to the same fault object based on prior geological knowledge. Godefroy et al. (2019) focused on the mathematical formalism of structural information in the form of graphs, as well as the definition of geological rules and the automated detection of major structures. This

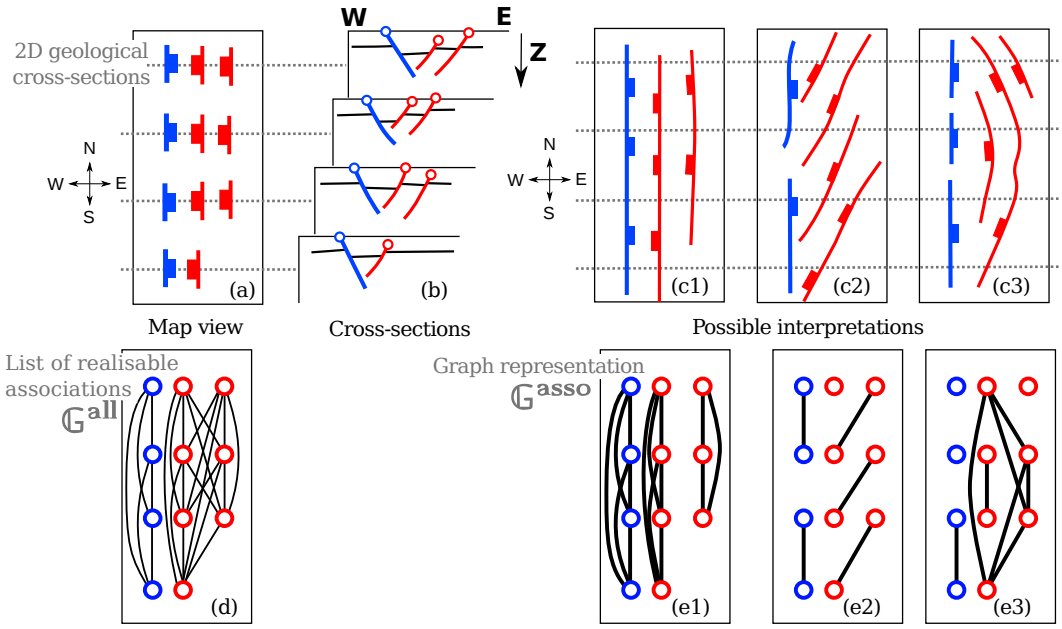

Figure 1. Associating labeled pieces of fault evidence (red: east-dipping and blue: west-dipping) interpreted (a) in map view or (b) on two-dimensional seismic lines is an under-constrained problem. (c_1, c_2 and c_3) Several structural interpretations are possible (Modified from Freeman et al., 1990; Godefroy et al., 2019). (d) In $\mathbb{G}^{\mathbf{all}}$, the labeled nodes (i.e., the pieces of fault evidence) are linked by an edge if they may be part of the same fault. (e_1, e_2 and e_3) Plausible interpretations are represented by an association graph where the edges link pieces of fault evidence interpreted as belonging to the same fault.

manuscript extends this work to the automated generation of multiple interpretation scenarios, which may help to assess fault-related structural uncertainty.

We propose a stochastic graph decomposition algorithm to automate the generation of several possible fault scenarios representing fault data and some structural knowledge (or, equivalently, graphs $\mathbb{G}^{\mathbf{asso}}$) from an input graph $\mathbb{G}^{\mathbf{all}}$ representing all the potential associations (Section 2). A graph data structure enables the generation of millions of alternative models. To study the properties of the model space sampled by this algorithm, we consider a reference model built from high-resolution seismic data, offshore Brazil, and extract from this reference model several sparse data sets of variable density. We combine several likelihood criteria translating varying degrees of geological knowledge to check the consistency of the method (Section 3). The results obtained are used to discuss how to further address some longstanding challenges for integrating data and knowledge and better understand brittle structures in the Earth's crust while addressing uncertainty (Section 4).

1 Structural uncertainty: state of the art

Structural uncertainty has received much attention in subsurface resource exploration and exploitation (Hollund et al., 2002; Richards et al., 2015; Rivenæs et al., 2005; Seiler et al., 2010), waste disposal (Mann, 1993; Schneeberger et al., 2017), environmental engineering (Rosenbaum & Culshaw, 2003), or civil engineering works (Zhu et al., 2003).

To characterize structural uncertainty, one may ask a population of geologists to interpret a particular data set (e.g., Bond et al., 2007; Schaaf & Bond, 2019). However, interpreting a subsurface data set in three dimensions commonly takes up to several months, so this strategy is difficult to generalize. Alternatively, the promise of 3D stochastic modeling is to use computing power to assess structural uncertainty. Stochastic structural modeling has already been proposed to generate several scenarios while taking account of seismic image quality and faults below seismic resolution (Aydin & Caers, 2017; Hollund et al., 2002; Holden et al., 2003; Irving et al., 2010; Julio et al., 2015a, 2015b; Lecour et al., 2001); uncertainty related to reflection seismic acquisition and processing (Osypov et al., 2013; Thore et al., 2002); geological field measurement uncertainty (Jessell et al., 2014; Lindsay et al., 2012; Pakyuz-Charrier et al., 2019; Wellmann et al., 2014); structural parameters for folding (Grose et al., 2019, 2018); and observation gaps (Aydin & Caers, 2017; Cherpeau et al., 2010b; Cherpeau & Caumon, 2015; Holden et al., 2003). Considering several structural interpretations has also proved useful to propagate uncertainties to subsurface flow problems (Julio et al., 2015b), to rank structural models against physical data and ultimately to falsify some of the interpretations using a Bayesian approach (Cherpeau et al., 2012; de la Varga & Wellmann, 2016; Irakarama et al., 2019; Seiler et al., 2010; Suzuki et al., 2008; Wellmann et al., 2014).

Generating realistic structural interpretation scenarios on a computer calls for the formulation of geological concepts in numerical terms, and for efficient computational techniques to explore the possibility space. Statistical point processes provide a general mathematical framework for this (Holden et al., 2003). As tectonic history places specific constraints on fault networks in terms of orientation and truncation patterns, it is possible to represent each fault surface as a level set and to sequentially simulate fault sets to reproduce specific statistics for each fault set, while enforcing abutting relationships between the simulated faults (Aydin & Caers, 2017; Cherpeau et al., 2010b, 2012; Cherpeau & Caumon, 2015). For honoring spatial fault data, Aydin and Caers (2017) use an extended Metropolis sampler which, at each stage of the simulation, adds, removes, or modifies a fault object. This sampler has theoret-

ical convergence properties, but simulating fault networks in the presence of a large number of fault data remains computationally challenging. Therefore, Cherpeau et al. (2010b, 2012) and Cherpeau and Caumon (2015) propose a parsimonious method which anchors the first simulated faults to the available evidence, before simulating unseen fault objects. All these iterative stochastic fault models are difficult to use in practice, because of the combinatorial complexity of the problem (Godefroy et al., 2019; Julio, 2015), of the difficulty to integrate geological, kinematical and mechanical concepts into the stochastic model (Godefroy et al., 2017; Laurent et al., 2013; Nicol et al., 2020; Røe et al., 2014; Rotevatn et al., 2018), and of the geometric challenges to robustly build such three-dimensional structural models (e.g., due to meshing issues, see Anquez et al., 2019; Zehner et al., 2015).

In this paper, we thus only focus on the problem of interpreting and associating available pieces of fault evidence. The graph framework of Godefroy et al. (2019) is easier to manipulate than a 3D structural geological model, making it a good candidate for developing stochastic modeling methods.

2 Multi-scenario interpretations using graph decomposition

In the graph framework proposed by Godefroy et al. (2019), the simulation of fault network scenarios amounts to decomposing the list of realisable associations $\mathbb{G}^{\mathbf{all}}$ into one or several association scenarios $\mathbb{G}^{\mathbf{asso}}$. An association graph is composed of disjoint and fully connected sets of nodes, each of these sets corresponding to fault surfaces. The graph is fully connected, meaning that if A and B belong to the same fault, and B and C also do, then A and C belong to that same fault. Disjoint implies that an observation can not belong to two faults at once. In the graph theory terminology, such a subset of nodes, that are all connected to one another, is referred to as a *clique*. In graph theory, cliques that cannot be enlarged without adding new edges are *maximal cliques*. The simulation of a fault network is thus equivalent to generating random decompositions of $\mathbb{G}^{\mathbf{all}}$ into a set of cliques. Several graph clustering methods are available in the literature (e.g., Schaeffer, 2007). These algorithms generally provide one single decomposition and are thus not directly applicable to uncertainty assessment by stochastic simulation. From $\mathbb{G}^{\mathbf{all}}$, a set of maximal cliques per fault family are thus computed. Cliques are then drawn from these maximal cliques to obtain an interpretation scenario.

2.1 Accounting for geological knowledge

Before explaining the proposed decomposition algorithm, we first describe how some geological information can be translated into the language of the graph framework, and the existing limitations of that framework.

2.1.1 Fault families

Regional geological knowledge arising from the description of tectonic phases through time is often used to group fault and fracture into families, for instance considering the apparent orientations in map view or in cross-section (Nixon et al., 2014). Stereograms displaying fault orientations are probably the most common tools to define fault families (Ramsay & Huber, 1987). In conjugate fault networks, or in horst and graben structure, fault dip also can be used to define two fault families (e.g., Figure 1). Interactions with other structures are also used; for example, whether faults are eroded or cross-cut by a particular erosion surface helps determining families of

faults active at different geological times. Interactions with other faults (Henza et al., 2011), syn-sedimentary structures or salt diapirs (Tvedt et al., 2016) also guide structural interpretation. Data obtained by inversion of focal mechanisms (Álvarez-Gómez, 2019) and mineralization observed within the fault core during mapping can also be used to cluster the pieces of evidence into distinct families.

As the number of fault data for a given area can be very large, we propose to use the concept of fault families to reduce the number of possible associations. These geological concepts can be translated in terms of statistical descriptions for fault and fracture families. In this context, a probability to belong to a particular fault family can be attached to each piece of fault evidence. Family rules $R_{\varphi}^{\mathbf{fam}}(\mathtt{v}_i)$ thus quantify the likelihood that a fault data $\mathtt{v}_i$ belongs to the given family denoted by the index φ (Godefroy et al., 2019). Family rules attach to each piece of evidence a number between 0 (if $\mathtt{v}_i$ cannot belong to the fault family φ) and 1 (if it is highly likely that $\mathtt{v}_i$ belongs to the fault family φ). This score is attributed by comparing semantic information about a piece of fault evidence, stored in the form of node labels, and general prior information about a given fault family. The list of realisable associations can thus be decomposed into several disjoint association subgraphs:

$$\mathbb{G}^{\mathbf{all}} = \bigcup_{\varphi=1}^{\varphi=n} \mathbb{G}_{\varphi}^{\mathbf{all}}, \tag{1}$$

where n is the number of fault families ($n = 2$ in Figure 1.e). In the list of realisable associations $\mathbb{G}^{\mathbf{all}}$, the decomposition into family subgraphs $\mathbb{G}_{\varphi}^{\mathbf{all}}$ may also be based on family marks attached to each piece of fault evidence (Figure 1.d).

2.1.2 Associating pieces of evidence

To determine which pieces of evidence belong to the same fault, geological concepts including expected fault orientation, size, kinematic or corrugation are used. The broad literature on scaling laws (Bonnet et al., 2001; Torabi & Berg, 2011) can also help determining whether two observations separated by a given distance may belong to the same fault. Kinematic analysis methods, such as the analysis of the throw distribution along strike, are also used during structural interpretation (Nixon et al., 2014; Tvedt et al., 2016). Well or seismic data on which no fault can be interpreted may also help interpreting that two pieces of evidence cannot belong to the same fault.

To include such a geological knowledge in the sampler, an association rule $R_{\varphi}^{\mathbf{assoc}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j)$ quantifies the likelihood that two pieces of evidence ($\mathtt{v}_i$ and $\mathtt{v}_j$) of the same family (φ) belong to the same fault (Godefroy et al., 2019). Association rules are defined from general structural concepts about faults and from some geometric characteristics associated with each fault family. An association rule $R_{\varphi}^{\mathbf{assoc}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j)$ returns a number between 0 ($\mathtt{v}_i$ and $\mathtt{v}_j$ cannot belong to the same fault of the family φ) and 1 (if both fault data are likely to belong to the same fault).

The proposed framework helps formalizing structural interpretation but currently suffers from three main limitations. First, the rules consider only two pieces of evidence at a time. This restricts the integration of geological methods such as throw map analyzis (Baudon & Cartwright, 2008). Second, the rules are applied directly on the observations rather than on a structural model. While this makes it possible to screen scenarios in an early stage of structural interpretations, validation methods such as structural restoration require a full 3D model. Finally, fault intersections and branching are not currently considered. Limitations and way forwards are discussed further in Section 4.3.

2.2 Size of the search space

From a combinatorial standpoint, the total number of association scenarios of n fault data is equal to the Bell number B_n, which correspond to the number of partitions of a complete graph (Godefroy et al., 2019). When rules and families remove edges in $\mathbb{G}^{\mathbf{all}}$, the graph becomes incomplete and the Bell number significantly overestimates the number of possible association scenarios. For example, without making any assumption or geometric consideration, the total number of possible scenarios for associating the 11 fault data in Figure 1 would be equal to $B_{11} = 678,570$. The definition of two disjoint families containing 4 and 7 pieces of evidence reduces this number to $B_4 \times B_7 = 13,155$.

In the general case where $\mathbb{G}_{\varphi}^{\mathbf{all}}$ are not complete subgraphs, the numbers of possible scenarios decrease dramatically. This impacts the comparison of results obtained with variable numbers of data, as the size of the search space cannot be evaluated without solving the highly challenging problem of explicitly listing all the possible configurations (Knuth, 2005).

2.3 Sampling by stochastic graph partitioning

To sample some of the possible scenarios, we propose a hierarchical method relying on sets of nodes that are all connected (cliques) in the subgraph of realisable associations of at least one family $\mathbb{G}_{\varphi}^{\mathbf{all}}$. Maximal cliques are used at the beginning of the process in order to mimic a manual interpretation where the geologist starts by interpreting the major structures.

Choice of parameters: based on the available data, pieces of fault evidence are digitized and represented as graph nodes. The method has been developed within the SKUA-GOCAD environment (Emerson, 2018) and takes advantage of the available data structures, so every graph node can be attached to a set of points, lines, or triangulated surfaces describing fault evidence geometry. Information attached to these pieces of evidence, such as throw or orientation, are stored as object properties within SKUA-GOCAD. For each fault family, a family rule and an association rule expressing prior geological knowledge are defined (Figure 2, step 1).

Creation and segmentation of the set of realisable association: the family and association rules are used to compute a graph of all realisable associations $\mathbb{G}_{\varphi}^{\mathbf{all}}$ for each family φ (Figure 2, step 2.1). A wide range of formulas can be used to compute the likelihoods according to the defined rules. Each edge $\mathtt{e}^{\varphi}(\mathtt{v}_i, \mathtt{v}_j)$ of $\mathbb{G}_{\varphi}^{\mathbf{all}}$ (linking two pieces of evidence $\mathtt{v}_i$ and $\mathtt{v}_j$) carries an association likelihood $L_{\varphi}^{\mathbf{all}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j)$ for each family φ:

$$L_{\varphi}^{\mathbf{all}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j) = R_{\varphi}^{\mathbf{fam}}(\mathtt{v}_i)\, R_{\varphi}^{\mathbf{fam}}(\mathtt{v}_j)\, R_{\varphi}^{\mathbf{assoc}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j). \tag{2}$$

For simplicity, this choice assumes that the events '$\mathtt{v}_i$ belongs to family φ', '$\mathtt{v}_j$ belongs to family φ' and '$\mathtt{v}_i$ and $\mathtt{v}_j$ belongs the same fault' are independent. It ensures that $L_{\varphi}^{\mathbf{all}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j) = 0$ if either $R_{\varphi}^{\mathbf{fam}}(\mathtt{v}_i) = 0$, $R_{\varphi}^{\mathbf{fam}}(\mathtt{v}_j) = 0$ (e.g., the dip of either pieces of fault evidence does not correspond to the dip of the family φ), or $R_{\varphi}^{\mathbf{assoc}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j) = 0$ (e.g., the locations of $\mathtt{v}_i$ and $\mathtt{v}_j$ are incompatible with the orientation of faults belonging to the family ϕ). The edges where $L_{\varphi}^{\mathbf{all}}(\mathtt{v}_i \leftrightarrow \mathtt{v}_j)$ is null are, therefore, removed from $\mathbb{G}_{\varphi}^{\mathbf{all}}$. This deletion has a strong impact on the simulation results, as the corresponding associations are not considered later, but it significantly reduces the number of acceptable association scenarios (Section 2.2). The major possible structures are then listed using the Bron-Kerbosch algorithm (Bron & Kerbosch, 1973) which finds the maximal cliques in $\mathbb{G}_{\varphi}^{\mathbf{all}}$ (Figure 2, step 2.2).

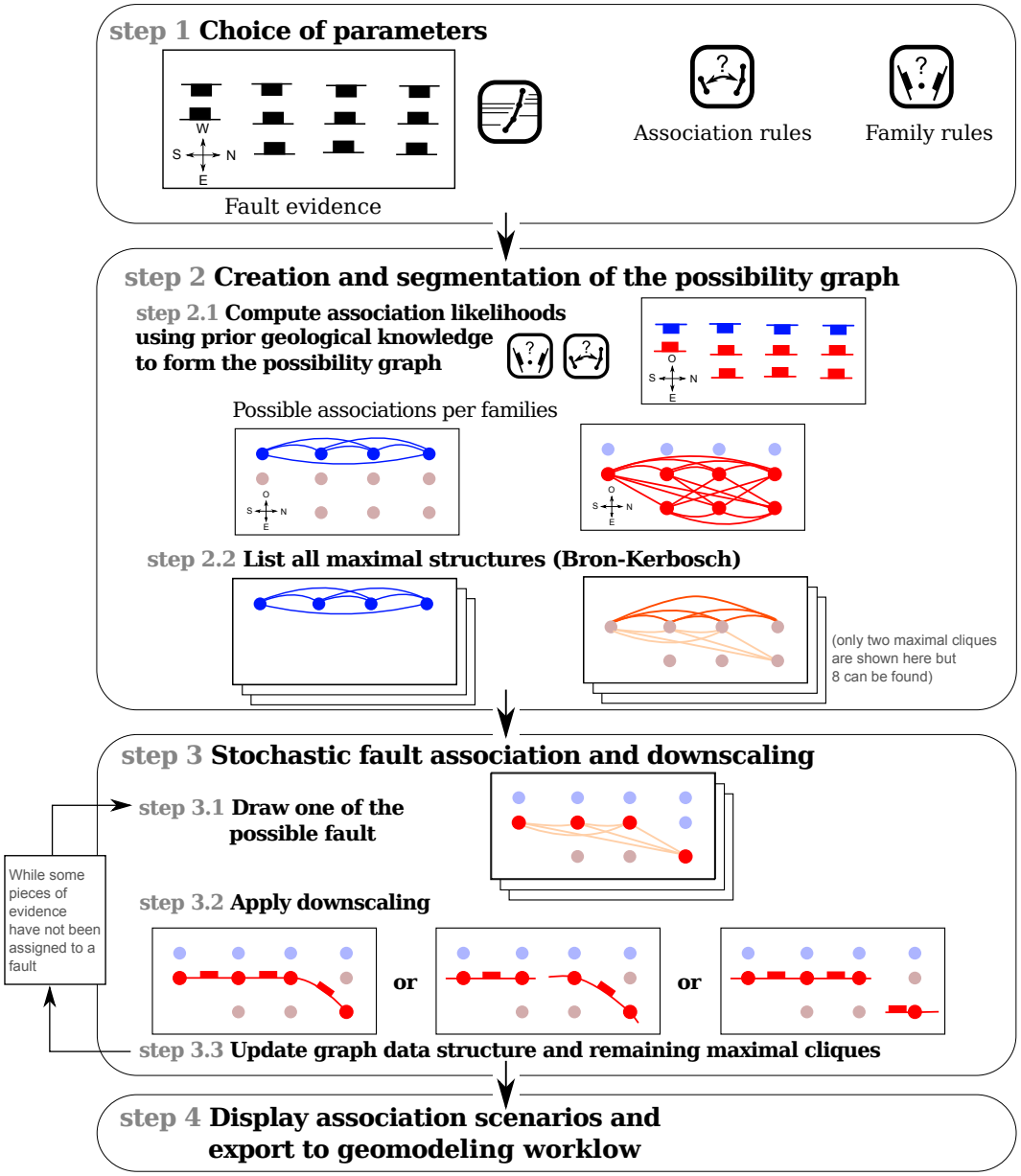

Figure 2. Sequential stochastic algorithm interpreting the available fault data as distinct cliques in $\mathbb{G}_{\varphi}^{\mathbf{all}}$. (step 1) The algorithm requires input fault interpretations and a set of interpretation rules. (step 2.1) Family and association rules are used to compute the graphs $\mathbb{G}_{\varphi}^{\mathbf{all}}$ of all possible associations for each family φ. (step 2.2) The potential major structures (maximal cliques) are detected. (step 3) Iteratively sample some fault objects associating a set of data and update the graph $\mathbb{G}_{\varphi}^{\mathbf{all}}$ and its maximal cliques. (step 4) When all the pieces of evidence have been assigned to fault surfaces, the association scenario can be displayed or used to interpolate or simulate the fault surface geometry.

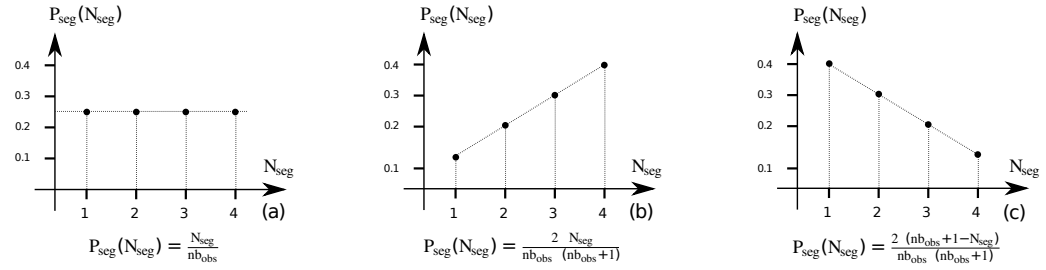

Figure 3. Three probability mass functions that can be used to draw the number of segments during the segmentation step: (a) uniform, (b) increasing, and (c) decreasing.

Stochastic fault association: cliques are randomly and sequentially drawn and removed from the remaining realisable associations until each fault evidence has been assigned to a fault. Several strategies can be defined and chosen for this sequential random selection of faults. To mimic the interpretation process by experts, who tend to first focus on the major structures (e.g., Lines & Newrick, 2004), we propose to preferentially select large and overall likely faults before selecting small and unlikely faults. At each step, the sampling probability of a clique $\mathbb{F} = \{\mathtt{v}_i, ..., \mathtt{v}_j\}$ depends on the number of nodes $|\mathbb{F}|$ and on the mean association likelihood $\overline{L_\varphi^{\mathbf{all}}(\mathbb{F})}$ as

$$P_{draw\ struct}(\mathbb{F}) = \frac{\overline{L_\varphi^{\mathbf{all}}(\mathbb{F})}\ |\mathbb{F}|^{\alpha_{draw}}}{\sum_{\mathbb{F}}^{\mathtt{all\ cliques}} \overline{L_\varphi^{\mathbf{all}}(\mathbb{F})}\ |\mathbb{F}|^{\alpha_{draw}}}, \tag{3}$$

where α_{draw} is used to weight the number of fault evidences in the clique (the structure containing more nodes are more likely to be drawn when α_{draw} increases, see sensitivity study in Appendix 6.2). Other selection strategies using, for example, the distance separating the pieces of evidence or their sizes could also be used to create large structures.

Stochastic fault segmentation The maximal clique listing is a way to start the interpretation in a parsimonious manner by looking at the major potential structures. However, these potential major structures can be made of several faults or several segments possibly linked by relay zones (e.g., Ferrill et al., 1999; Julio et al., 2015a; Manighetti et al., 2015; Peacock & Sanderson, 1991). These potential fault segments should be considered to completely explore the uncertainty space. For this, we propose a simple procedure to split a fault $\mathbb{F}$ made of $|\mathbb{F}|$ pieces of fault evidence into N_{seg} fault segments; this strategy is called *downscaling* in Julio et al. (2015a, 2015b). For simplicity, the number N_{seg} is drawn randomly in this paper between 1 (no segmentation) and $|\mathbb{F}|$ (each piece of evidence explains one individual fault segment). This random selection relies on a discrete probability distribution called P_{seg}, which can be either uniform, linearly decreasing or increasing (Figure 3.a, b, c, respectively). A sensitivity analysis is presented in Appendix 6.2 to show how the choice of P_{seg} impacts the total number of detected structures.

Fault surface modeling: the outcome of the simulation process is a set of association scenarios, each being represented by an association graph (Figure 4.a). Converting the graphs into fault network models (Figure 2, step 4) is beyond the scope of this paper. While for simple synthetic cases, fault geometry can be approximated using geometrical criteria (for example by ellipses computed from

the data, Figure 4.b), determining the tip line and fault geometry from the data is non trivial and also prone to uncertainties in the general case.

In the above simulation method, faults are simulated independently, and the possibility of interactions is not considered. Even though branch lines are an ubiquitous feature of fault networks, branch lines are difficult to map directly from subsurface data (Yielding, 2016). In modeling workflows, branch lines are thus interpreted from the nearby fault sticks. For this reason, we consider that each piece of fault evidence can belong only to one fault surface, and the nodes corresponding to a selected clique are not considered in further simulation steps (Figure 2, step 3.3).

When fault data are interpreted along parallel two-dimensional seismic lines, an extra constraint can be added to prevent fault-fault intersections. The obtained associations are then free of branch lineswhich. In the sequential fault association framework, after each downscaling step (step 3.2), the edges crossing the simulated fault are removed from the remaining cliques while updating the data structure (step 3.3). A similar graph problem is solved using the Dynamic Time Warping algorithm (Levenshtein, 1966) in order to correlate stratigraphic data along wells (Edwards et al., 2018; Lallier et al., 2013; Smith & Waterman, 1980). Note that the proposed strategy to avoid intersections is applicable only in the presence of fault interpretations made on parallel sections. Extending this constraint to irregularly sampled sparse data could be achieved, for example, by using visibility criteria (Holgate et al., 2015).

3 Application to sparse data from Santos basin, offshore Brazil

We applied the proposed stochastic fault network simulation method on a natural example of faulted structures located in the Santos basin. The Santos Basin is located offshore SE Brazil and is one of several salt-bearing passive-margin basins flanking the South Atlantic (Jackson et al., 2015).

The Santos Basin formed as a rift during the Early Cretaceous when the South Atlantic began to open (e.g., Meisling et al., 2001). Grabens and half grabens were filled by largely Barremian, fluvial-lacustrine deposits, which are overlain by an early-to-middle Aptian, carbonate-dominated succession (Jackson et al., 2015). During the late Aptian, a thick (up to 2.6 km) salt-rich succession was deposited (see Tvedt et al. (2016) and references therein). During the Albian, marine conditions established in the Santos Basin, leading to the deposition of carbonate-dominated succession (Ithanhaem Formation). An abrupt increase of the water depth in Santos Basin, in Cenomanian–Turonian times, drown the Albian carbonate system as a fine-grained, clastic-dominated succession accumulated (Itajai-Acu Formation).

The growth of the observed faults was activated by the mobilization of the underlying salt-rich unit (Ariri Formation) (Jackson et al., 2015; Tvedt et al., 2016)., as thin-skinned normal faulting systems accommodated the overburden stretching above the mobile salt. Previous kinematic analysis (Tvedt et al., 2016) showed that the faults grew from Albian to Miocene and from Oligocene to present within Albian Carbonates (Ithanhaem Formation) and within Cenomanian to recent fine-grained clastics (Itajai-Acu and Marambia formations),

From a time-migrated three-dimensional seismic image (see sample section on Figure 5.a), we selected a densely faulted area where we interpreted 27 fault surfaces (see Figure 5.b and Godefroy et al. (2017)). The modeled zone covers an area of 6 km per 6 km and is free of branch lines, simplifying the application of our methodology. The lengths of the observed faults range from a few hundred meters to 3.6 km.

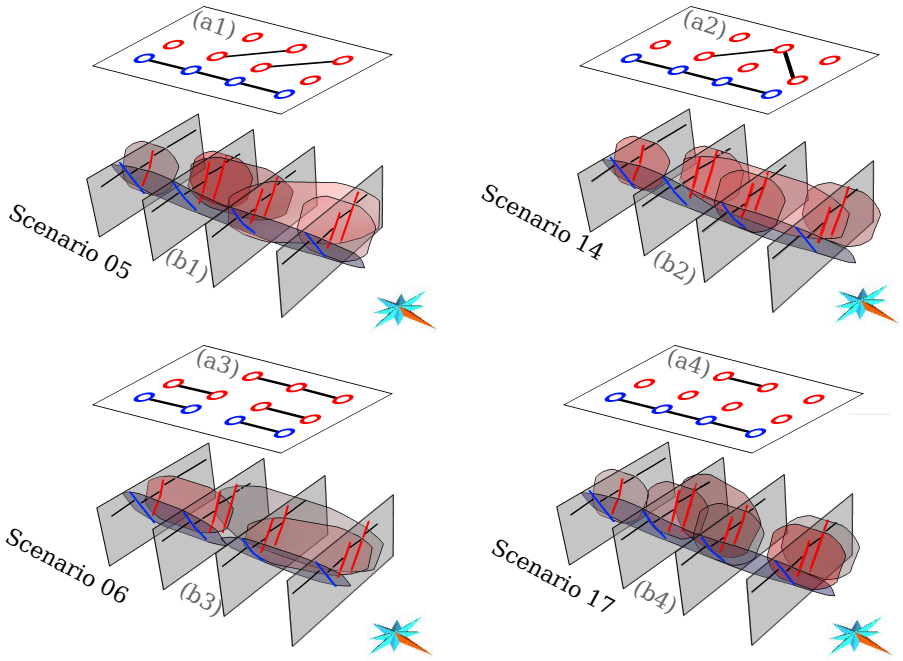

Figure 4. (a) The generated interpretation scenarios are represented by association graphs. (b) For each clique, a fault surface can be interpolated using the geometry of the fault evidence. Data and generated models can be visualized at: `gabrielgodefroy.github.io/StochasticInterpData/Fig4/Fig4.html`

Given the excellent quality of the seismic image, there is very limited structural uncertainty in this interpretation, which is used below as the reference interpretation. From this reference model, we extracted sparse fault data along several parallel two-dimensional sections to emulate the case of the same area being imaged by two-dimensional seismic lines (Figure 5.c,d).

In the remainder of Section 3, we propose numerical experiments to evaluate the consistency of the model space sampled by the proposed simulation method from these incomplete data sets. Intuitively, a consistent sampling method should, when appropriately parameterized, retrieve the reference association with the maximum frequency. Also, the likelihood to retrieve the reference association should increase as more data and correct informative rules are used. However, in practice, the rules may be biased because of preferential sampling or wrong analog knowledge, so we will also check for the impact of using biased rules on the ability of the method to find the correct association. Finally, in a consistent sampling, the spread of the samples around the reference should reduce when more information becomes available. However, checking for the latter property is difficult, as the dimension of the problem changes when the number of observations changes. Therefore, we first study how the structure of the association problem changes with the number of fault data and the degree of information brought by geological rules.

3.1 Synthetic two-dimensional lines

To quantify the role of a particular geological concept in reducing structural uncertainty, we now consider several interpretation rules applied to several data sets of increasing density extracted from the reference model. The quality of the 3D seismic image used to build the reference model enables to determine whether the reference association can be retrieved, and to study the influence of chosen geological rules and algorithm parameters on the quality of generated interpretations.

The multi-scenario association strategy was applied on fault evidence extracted along 3, 4, 5, and 6 cross-sections (see Figure 5 and online material). The availability of a trustworthy reference fault network model allows us to study the proposed stochastic interpretation methodology using a set of restrictive geological rules consistent with this reference model. Such an ideal case is unrealistic in actual sparse data settings but it enables us to test how the rule choices impact structural interpretation by deleting and modifying some of the rules.

3.2 Association rules

In the considered data set, faults have approximately a north/south strike and can be grouped into two fault families: east- and west-dipping faults. Two fault family rules are defined based on the dip direction:

$$R^{\mathbf{fam}}_{\varphi_1}(\mathbb{v}_i) = \begin{cases} 1, & \text{if } \mathbb{v}_i \text{ dips towards the West,} \\ 0, & \text{otherwise.} \end{cases}$$

and

$$R^{\mathbf{fam}}_{\varphi_2}(\mathbb{v}_i) = \begin{cases} 1, & \text{if } \mathbb{v}_i \text{ dips towards the East,} \\ 0, & \text{otherwise.} \end{cases}$$

As seismic lines are oriented east/west, the slope of fault interpretations completely determines the family, so there is no uncertainty about which family each piece

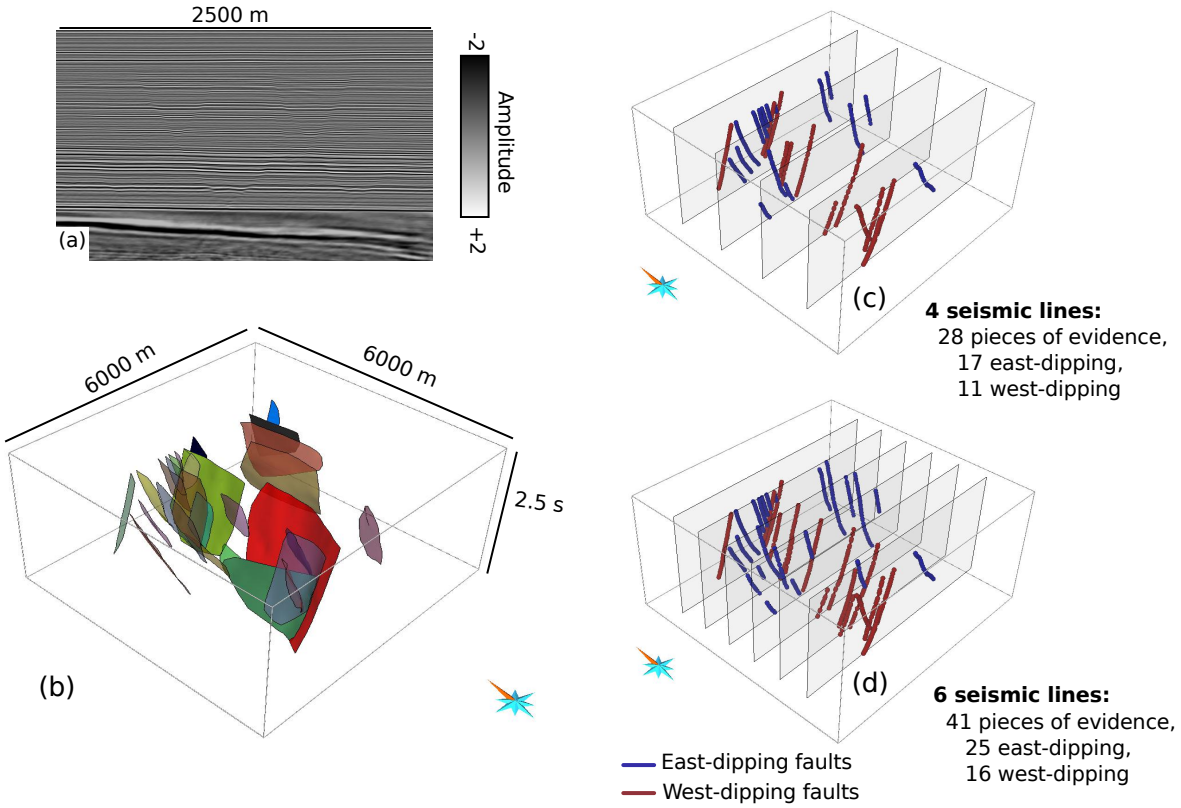

Figure 5. Reference structural model located in the Santos Basin, offshore Brazil. (a) Available reflection seismic data (courtesy of PGS). (b) Reference fault network. (c, d) Generated interpreted synthetic parallel two-dimensional seismic lines. An interactive three-dimensional model is provided as supplementary material, in the form of an HTML page.

of evidence belongs to. East- and west-dipping are respectively identified in blue and red on Figure 5.

To evaluate the likelihood of associating two fault traces of the same family, we first use an association rule that restricts the strike of the generated faults to be between $N330$ ($strike^{min}$) and $N015$ ($strike^{max}$) for both families:

$$R^{orient}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = \begin{cases} 1, & \text{if the strike between } \mathbb{v}_i \text{ and } \mathbb{v}_j \text{ is between } strike^{min} \text{ and } strike^{max}, \\ 0, & \text{otherwise.} \end{cases}$$

Additionally, a uniform association distance rule is created to prevent the interpreted fault lengths from being longest than $dist_{max} = 3.5$ km, the maximum fault length observed in the reference model:

$$R^{dist}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = max(\frac{1 - dist(\mathbb{v}_i \leftrightarrow \mathbb{v}_j)}{dist_{max}}, 0).$$

Finally, both rules are combined using:

$$R^{\mathbf{assoc}}_{\varphi}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = R^{dist}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) R^{orient}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j).$$

The above association rules are simple and rely only on one orientation and one distance criterion; a structural geologist would also use (at least) the length of the fault sticks, the fault throw, and the geometric relations between the different observations.

To assess the impact of conceptual information on the structure of the list of realizable associations (Figure 6), we first consider each association rule separately. As expected, restrictive rules decreases the number of edges in $\mathbb{G}^{\mathbf{all}}_{\varphi}$, hence the number of possible association scenarios for each family (see Section 2.2). In the interpreted area of interest, faults do not intersect each other, so intersections are also forbidden during the simulations. In spite of these rules, going from 4 to 6 seismic sections increases both the number of graph nodes (from 28 to 41, see Table 1) and graph edges (from 378 to 820), making it more difficult to explore the search space and to find the reference fault configuration.

3.3 Evolution of the number of possible scenarios

The proposed sampling method may generate the same fault scenario several times. To assess whether the sampler has converged, a common strategy consists in generating models until the number of *distinct* scenarios stabilizes (Pakyuz-Charrier et al., 2019; Thiele et al., 2016). For this, we use the metric $N_{diff}(l, m)$ which counts the number of differences between any two realizations $\mathbb{G}^{\mathbf{asso}}_{l}$ and $\mathbb{G}^{\mathbf{asso}}_{m}$. N_{diff} is a special case of graph edit distance (Sanfeliu & Fu, 1983), in which the only edit operations are edge insertion and deletion:

$$N_{diff}(l, m) = \sum_{\mathbb{e} \in \mathbf{edges}} d_{l,m}(\mathbb{e}) \tag{4}$$

where

$$d_{l,m}(\mathbb{e}) = \begin{cases} 0 & \text{if the edge } \mathbb{e} \text{ is either present or missing in both } \mathbb{G}^{\mathbf{asso}}_{l} \text{ and } \mathbb{G}^{\mathbf{asso}}_{m}, \\ 1 & \text{if not.} \end{cases} \tag{5}$$

Table 1. Number of pieces of fault evidences, corresponding simulation times and Bell number for different number of synthetic seismic lines. Simulations were performed using all rules described in Section 3.2. Simulations were carried out on a PC with an Intel Xeon CPU E5-2650 v3 @ 2.30GHz with 64GB of RAM; the code is not parallelized.

# seismic lines	# data...	... in φ_1	... in φ_2	# edges in ref. model	Bell numbers	Run time (5×10^7 real.)
3	25	16	9	2	$B_{25} = 4.6 \times 10^{18}$	03h49m
4	28	17	11	5	$B_{28} = 6.1 \times 10^{21}$	03h51m
5	35	22	13	13	$B_{35} = 2.8 \times 10^{29}$	05h30m
6	41	25	16	19	$B_{41} = 2.3 \times 10^{36}$	06h54m

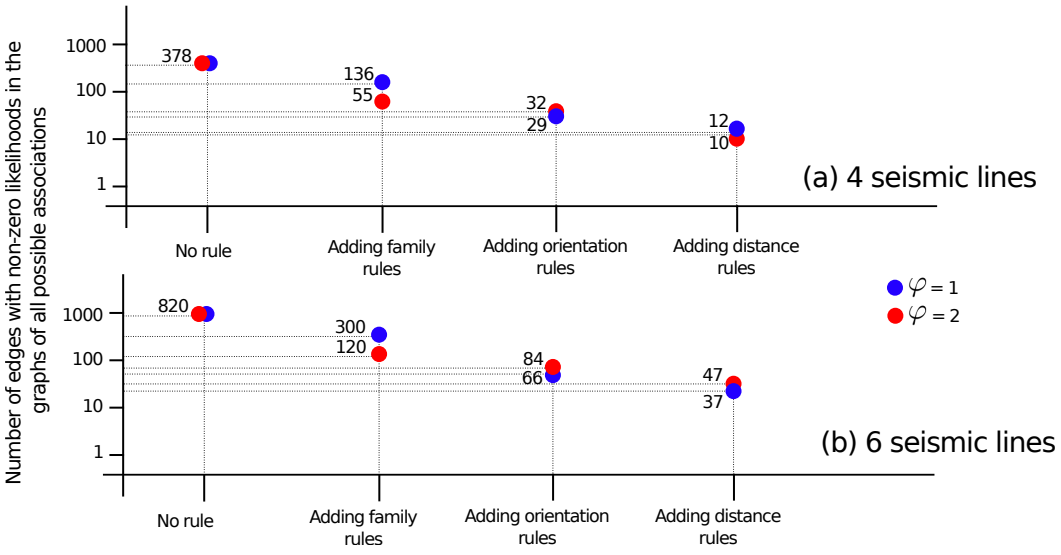

Figure 6. Number of edges per fault family (φ) in the graphs of all possible associations ($\mathbb{G}_{\varphi}^{\mathbf{all}}$) for evidence extracted along (a) 4 seismic lines, and (b) 6 seismic lines. The integration of geological rules (Appendix 6.1) reduces the density of the graphs of all possible associations $\mathbb{G}_{\varphi}^{\mathbf{all}}$. The lists of realizable associations can be interactively visualized here:

`gabrielgodefroy.github.io/StochasticInterpData/Fig6/html/Fig6.html`

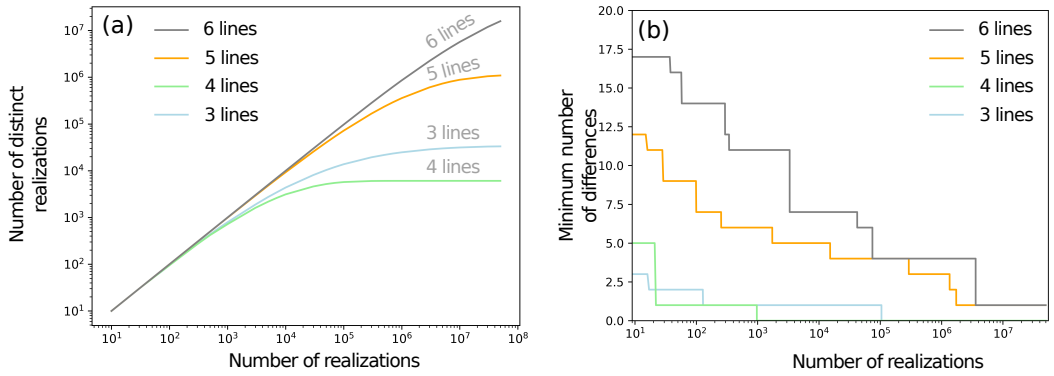

Figure 7. (a) Number of distinct association scenarios obtained from 5×10^7 realizations from 3, 4, 5, and 6 seismic lines. A plateau is not reached for the case with 5 and 6 seismic lines, highlighting the very large computational complexity of the problem. (b) Minimum number of differences to the reference found over the first realizations.

For computational reasons, 5×10^7 realizations were simulated from the fault interpretations extracted from 3, 4, 5, and 6 virtual seismic lines, using all the previously described rules (Figure 7.a). Simulations run in 3 to 7 hours (see Table 1 for details). At the beginning of the simulation, the sampling algorithm shows a near-optimal exploration efficiency as it generates only different realizations, whatever the data density.

For the case with 3 seismic lines, the plateau is not yet fully reached but for the case with 4 seismic lines, a plateau of 6072 distinct realizations is reached after 3×10^6 realizations. In both cases, the reference association is found 2026 times and 3164 times for 3 and 4 seismic lines, respectively. This can be explained by the different information content carried by these data sets (see Figure 6 of the supplementary material). When simulating from 5 and 6 seismic lines (35 and 41 fault data, respectively), the numbers of distinct scenarios is still significantly increasing after 5×10^7 realizations and the reference association is not found. In both cases, the best scenario produced has a number of differences N_{diff} equal to 1 (Figure 7.b).

This numerical experiment shows the difficulty of retrieving the reference association when the number of pieces of fault evidence is high, even if the chosen rules are informative and consistent. Indeed, as discussed in Section 2.2, the combinatorial complexity increases in a non-polynomial way with the number of nodes. As a consequence, the dimension of the search space to be explored by the algorithm is significantly higher when more data become available. Therefore, the reference model (and every realization) becomes diluted in the search space when the number of data increases. The difficulty to retrieve the reference model suggests that this dilution effect is not compensated by the information content carried by the association rules. In principle, the graph edit distance N_{diff} for different number of data should be normalized depending on the search space size and/or on the number of edges in the reference model (1).

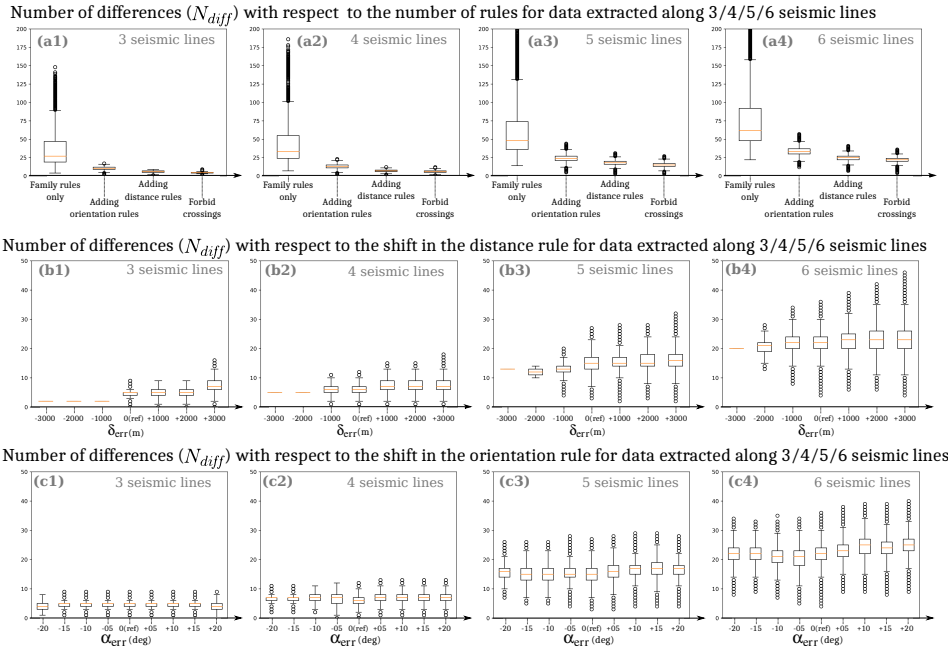

Figure 8. Box plots showing the minimum, mean and maximum number of differences (N_{diff}) between the simulated association graphs and the reference one. Graphs computed for sparse data extracted from 3 ($a_1 - c_1$) to 6 ($a_4 - c_4$) seismic lines. ($a_1 - a_4$) The integration of more geological rules reduces the range of possibilities, corresponding to lower density for the graphs of all geologically meaningful associations $\mathbb{G}_{\varphi}^{\mathbf{all}}$. Falsification of the distance rule ($b_1 - b_4$) and of the orientation rule ($c_1 - c_4$). Statistics computed over 5×10^5 realizations.

The experiment shows that the formalism and the chosen metric are favorable when relatively few pieces of fault evidence are available. This may seems counter-intuitive; however, when pieces of evidence are added, the difficulty to explore a larger search space increases in a non-polynomial way, as also observed by Edwards et al. (2018) for well correlation.

3.4 Influence of the chosen geological rules on simulation results

3.4.1 Impact of the number of rules

We tested the impact of the chosen numerical rules by successively running simulations with an increasing number of rules (orientation rule, distance between pieces of fault evidence, and forbidding fault intersections). 5×10^5 realizations were generated from digitized fault evidence extracted along the 3, 4, 5, and 6 seismic lines.

To analyze these results, we now consider the number of differences $N_{diff}(l, ref)$ between each simulated association $\mathbb{G}^{\mathbf{asso}^l}$ and the reference association $\mathbb{G}^{\mathbf{asso}}_{ref}$ interpreted from the full 3D seismic data set. When the simulations are run with more rules (Figure 8.$a_1 - a_4$), the minimum, mean, and maximum number of differences $N_{diff}(l, ref)$ consistently decrease. Moreover, we observe that rules interact with the data density on two respects. First, considering a single rule in sparse data settings (cases with 3 or 4 sections) yields realizations closer to the reference than considering more informative rules in denser data settings (cases with 5 or more sections). This can be explained by the non-convergence of the sampler for 5 and 6 sections, as shown in Figure 7.a. Second, for a given set of rules, the absolute number of differences from the reference increases with the number of data. This is a direct effect of the increasing complexity of the search space, and may also be explained by the non-convergence of the sampler with 5×10^7 realizations for more than 5 seismic lines (Figure 8.a_4).

It would be interesting to weight these distributions by a relative likelihood for each particular scenario using the correlation rules. If appropriately chosen, the rules should then give a larger weight to the scenarios closer to the reference. However, computing such a relative likelihood faces again a normalization challenge, as the number of graph edges is generally different for each realization.

3.4.2 Deliberately selecting biased rules

In practical geological studies, it would be difficult to come up with appropriate parameters for the association rules. Therefore, we tested the impact of choosing erroneous geological concepts during structural interpretation: the orientations defining the associations rules were shifted by an angle α_{err} and the distance $dist_{max}$ defining the distance association rule was offset by δ_{err}. The mathematical expressions for the falsified rules are given in Appendix 6.1.

We study the mean number of differences to the reference association (over 5×10^5 realizations) according to these inappropriate choices (Figure 8.b,c). Strongly biased rules make it impossible to retrieve the reference association from fault evidence sampled along 3, 4, 5 or 6 seismic lines (Figure 8.b_1, b_2, c_1, c_2, b_3, b_4 and c_3, c_4, respectively). When simulating interpretation scenarios from data extracted along 5 or 6 seismic sections, the reference scenario is never retrieved.

In the case of the distance rule, if δ_{err} is negative, no association is allowed, yielding a collapse of the ensemble of associations. Such a collapse clearly highlights an inconsistency between the rules and the spacing of the synthetic data. On the other hand, if the distance rule is more permissive than the reference one (i.e., δ_{err} is positive), then the minimum, mean, and maximum numbers of error increase for all number of seismic lines.

The effect of changing the orientation rule is not as dramatic, but for the cases with 5 and 6 seismic lines, the deviation from the reference rules leads to a slight increase of the mean number of errors.

4 Discussion and ways forward

4.1 On uncertainty and objectivity

A major difference between the proposed approach and the expert-based manual interpretation is the intrinsic ability of the former to generate several scenarios, whereas most of the latter tend to end up with one deterministic solution. A reason is

that interpretation exercises are taught as a deterministic activity in the vast majority of university courses: the general expectation, in surface or subsurface mapping, is to produce only the most likely scenario or model. This is shown even in experiments assessing the interpretation uncertainty, which ask a set of geoscientists to produce one interpretation each Bond2007, Bond2015b, Schaaf2019. Cognitive biases also explain the difficulty of one to work with multiple hypotheses Chamberlin1890, Wilson2019, which is a possible explanation for interpretation bias Bond2007. The advent of computer-based methods makes it easier to explore aleatory uncertainties by perturbing a reference model see Wellmann2018 and references therein, but addressing epistemic uncertainties is more challenging as it requires to formalize the geological concepts. The graph-based sampling method proposed in this paper clearly belongs to this latter class of methods.

A major difference between the proposed approach and the expert-based manual interpretation is the intrinsic ability of the former to generate scenarios, whereas most of the latter tends to end up with deterministic solution. Geoscience is a discipline where uncertainty is a key component (Frodeman, 1995) and geoscientists and interpreters are exposed to uncertainty from their first university courses. However, there is a large loss between how geologists see uncertainty and what is often produced as a result of their work. This is shown even in experiments assessing the interpretation uncertainty, which ask a set of geoscientists to produce one interpretation each (Bond et al., 2007; Bond, 2015; Schaaf & Bond, 2019). In most studies, a single deterministic structural model is produced, and structural uncertainty are communicated in the form of reports and diagrams. Cognitive biases also explain the difficulty of one to work with a large number of hypotheses in parallel (Chamberlin, 1890; Wilson et al., 2019). Cognitive biases have been demonstrated experimentally (Kahneman, 2011) and are a possible explanation for interpretation bias (Bond et al., 2007). The advent of computer-based methods makes it easier to explore aleatory uncertainties by perturbing a reference model (see Wellmann & Caumon, 2018, and references therein), but addressing epistemic uncertainties is more challenging as it requires to formalize the geological concepts. The graph-based sampling method proposed in this paper clearly belongs to this latter class of methods.

A second major difference with classical fault interpretation is that our approach relies on the interpreter to explicitly formulate and organize elementary association rules, whereas expert based knowledge is seldom explicitly described. The order by which we processed and combined the rules is driven in part by mathematical and algorithmic convenience, and we do not claim it to be fully objective. Indeed, geological interpretation always depends on the current state of knowledge and experience of an interpreter, on some model assumptions, and on the tools used during the interpretation (Bond et al., 2007; Chamberlin, 1890; Frodeman, 1995; Wellmann & Caumon, 2018). Nonetheless, as compared to classical expert-based interpretation, we see the general approach proposed in this paper as a step towards making the interpretation process more transparent and reproducible by expressing formally the geological concepts.

4.2 On graphs for structural uncertainty assessment

We see the graph-based method proposed in this paper as a possible way to complement or generalize structural uncertainty assessment method. It starts from existing observations and thus follows the same philosophy as perturbation strategies, which consider uncertainties in the location and/or orientation of observations (Lindsay et al., 2012; Pakyuz-Charrier et al., 2018; Wellmann et al., 2010) or modeled structures (Holden et al., 2003; Lecour et al., 2001; Røe et al., 2014). Our approach starts without any particular assumption about how to associate incomplete fault observations

together. Unlike previous iterative methods (Aydin & Caers, 2017; Cherpeau et al., 2010a; Cherpeau & Caumon, 2015), the potential major fault structures are processed in the early steps of our algorithm thanks to the maximal clique detection, assuming that largest faults are most likely to correspond to many graph nodes. This reproduces a classical interpretation process whereby geologists focus on largest structures before focusing on smaller objects (e.g., Lines & Newrick, 2004).

Approaches addressing topological fault network uncertainty have been proposed before, using mainly data-driven iterative simulation methods (Aydin & Caers, 2017; Cherpeau & Caumon, 2015; Julio et al., 2015b) or object simulation based on stochastic point processes (Hollund et al., 2002; Munthe et al., 1994). All data-driven approaches, including the method presented herein, can be seen as an efficient way to honor observations, which is a notably difficult and time-consuming process when the spacing between observations is smaller than the size of the simulated objects. Data-driven fault simulations are parsimonious as they only focus on explaining observations, but they can significantly under-estimate the number of faults in a given domain. For example, the Santos case study clearly shows that the number of simulated faults decreases when less data is used for the same area of interest (Figure 9). Therefore, we firmly believe that a stochastic point process should ultimately complement the proposed approach to simulate faults that are not directly supported by observations (Bonneau et al., 2016; Cherpeau et al., 2010b; Davy et al., 2013; Holden et al., 2003; Munthe et al., 1994; Stoyan & Gloaguen, 2011).

Another line of progress in the graph-based method concerns the management fault branch lines. Indeed, even though the chosen reference data set (Figure 5) is free of branch lines, fault branching and interactions are ubiquitous in faulted systems, and have been extensively studied to understand fault growth (Watterson, 1986; Walsh et al., 2002; Nixon et al., 2014; Nicol et al., 2020). The simulation process should be extended to represent how faults branch in the fault network while accounting for the chronology of the development of the successive fault families. A second oriented graph could represent fault branching. while preserving the spatial dependency of fault geometry, hierarchy and the fault abutting relationship (as in Aydin & Caers, 2017; Cherpeau et al., 2010b).

4.3 Are the produced interpretations "geologically realistic"?

Another major difference between expert-based interpretations and the proposed method concerns the amount of interpretive concepts used. Interpretation concepts typically arise from outcrop or subsurface analog data bases, laboratory and numerical models, or on an inference process applied directly to the data at hand. Those concepts are conveyed in the form of text, sketches and oral presentations. The proposed graph-based method calls for formally defining the rules according to each specific geological context. In this paper, we only tested relatively simple geological rules in order to assess the consistency of the sampling algorithm. If 3-D models were built, we could test also their validity by asking a population of geologist to visually inspect the models and rate their likelihood. Further studies are clearly needed to help geoscientists defining the numerical interpretation rules corresponding to the interpretation concepts. In the case of faulted structures, concepts include choosing fault surface orientations from analog data sets (Aydin & Caers, 2017), accounting for fault curvature and lateral extension, estimating the fault slip (Cherpeau & Caumon, 2015; Røe et al., 2014), and evaluating fault segmentation (Julio et al., 2015a; Manighetti et al., 2015; Manzocchi et al., 2019). Machine learning could also come into play in this process, either by inferring rule parameters or a posteriori assessing the likelihood of the various realizations produced by the sampling method. Training for these approaches

could be achieved on multiple manual interpretations (e.g., Schaaf & Bond, 2019) or on processed synthetic models (Wu et al., 2019).

To test the proposed sampler, we started by choosing an ideal case where the parameters of simple rules were calibrated directly on a reference model. We acknowledge that this is never the case in practice where no reference model exists, but this allowed us to compare the simulated scenario with a reference. Tests made in Section 3.4.2 suggest that, when rule parameters are chosen inappropriately, the simulated models can show significant bias, and we can expect that this observation would also hold for more complex rules. However, such an automatic method has no guarantee to produce the same results as several interpretations made by several experts. One reason is that the sampling method is likely to miss some important aspects of geological interpretation. On this regard, we see two main avenues for improvement:

- First, the methodology does not completely automate the three-dimensional structural modeling. Branch lines are not handled and a boundary representation of the subsurface is not generated. This makes it difficult to assess the likelihood of the generated fault networks using advanced structural analyzes such as global displacement analysis (Freeman et al., 2010) or structural restoration (Gratier & Guillier, 1993). For this, fault geometries should be modeled using explicit surfaces (Lecour et al., 2001; Røe et al., 2014) or implicit surfaces (Aydin & Caers, 2017; Cherpeau et al., 2010b). Then, the geological formations affected by the fault network should be modeled. Generating such geometries would also be useful to assess the impact of fault network uncertainty on resource assessment (Richards et al., 2015), to incorporate this source of uncertainty in geophysical inverse problems (Giraud et al., 2019; Ragon et al., 2018), or to consider the geological likelihood in a Bayesian inference problem (Caumon, 2010; de la Varga & Wellmann, 2016).
- Second, the graph formalism at this stage only considers pairwise associations but does not use the likelihood of associating several pieces of evidence at once. This calls for translating geological concepts into numerical rules which apply to all fault evidence at once. For example, one could consider the throw distribution along fault strike (as in Cherpeau & Caumon, 2015; Freeman et al., 1990) or statistical relationships between observed separations and fault size (e.g., Gillespie et al., 1992; Torabi & Berg, 2011). The association likelihoods should also be updated during the graph-based interpretation to account for the expected likelihood of fault-fault interactions.

4.4 Inverse problem and clustering of structural interpretations

Stochastic structural modeling enables the generation of large numbers of alternative scenarios (several millions in this work) which can be used as prior information in subsurface inverse problems (see Wellmann and Caumon (2018) and references therein). However, the computational times of focal mechanism inversion, flow simulation, or seismic forward modeling are often incompatible with more than hundreds of models. Furthermore, for a human being, it seems difficult to work with more than a few alternative scenarios deemed representative of the uncertainties. An effective way to address this problem is to use model clustering in model space (e.g., Suzuki et al., 2008) or in data space (e.g., Scheidt et al., 2018; Irakarama et al., 2019). A challenge, in both cases, comes from the redundancy of models sampled by a particular stochastic methods: indeed, simulation methods tend to generate many similar models in a priori likely regions of the search space. This redundancy is needed if the models also have a large posterior probability, but it can raise efficiency problems when the Bayesian updating is strong. The graph-based sampler described in this paper opens some avenues to make progress in this area. Indeed, maximal cliques are detected and

processed sequentially within the sampling algorithm. Therefore, a hierarchical clustering of structural scenarios could be generated by applying the method in a recursive manner. A possible and simplified outline of such a hierarchical sampling reads: 6.3.

5 Conclusions

The proposed graph-based framework helps interpreting alternative fault scenarios to account for the uncertainty arising while considering sparse fault sample. Prior geological knowledge is formalized using numerical rules. The mathematical format of the rules eases the communication of the geological concepts used during the interpretation and makes the structural interpretation process reproducible. Simple geometric rules are used in paper, and further developments on integrating more advanced geological concepts have been discussed.

Each scenario is represented by a graph. The automatic interpretation framework relies on the detection of the major possible structures in the graphs of all possible associations. This strategy mimics the behavior of an interpreter who would start by the larger structures. The use of a graph data structure, as compared to a full three-dimensional model, leads to a fast simulation process. This enables to perform sensitivity studies on the numerical rules and simulation parameters using sparse data extracted from a reference model. Converting the graph into a three-dimensional model remains a perspective of this framework.

The presented numerical experiments illustrate the difficulty in retrieving the correct association scenario from sparse data. Even if interpretation rules reduce the number of scenarios, it seems highly unlikely that a single interpretation is correct. When working with subsurface data, uncertainty is the norm and not the exception (Frodeman, 1995). These experiments also confirm that the simultaneous use of several coherent geological rules reduces the number of distinct simulated scenarios. The simulated models are closest (on average) to the reference model and rule falsification decreases the likelihood to find a scenario close to the reference one. These experiments formally confirmthat, in the interpretation setting defined in this paper, the importance of the prior geological knowledge during structural interpretation.

We also advocate for making geologists aware of structural uncertainties in the early stages of their training during geological education (Chamberlin, 1890) . Formalizing explicitly the interpretation concepts should ease their communication and limit interpretation biases.

Acknowledgments

This work was performed in the frame of the RING project (http://ring.georessources.univ-lorraine.fr/) at Université de Lorraine. We would like to thank for their support the industrial and academic sponsors of the RING-GOCAD Consortium managed by ASGA. Software corresponding to this paper is available to sponsors in the RING software package FaultMod2. We also acknowledge Paradigm for the SKUA-GOCAD Software and API. The authors are grateful to PGS Investigação Petrolífera Limitada and to Chris A.L. Jackson for providing the Santos Basin reflection seismic data. Readers can access the reference structural model and the generated cross-sections from : https://doi.org/10.17605/OSF.IO/MP97W.

6 Appendices

6.1 Rules expression

We give here the numerical formulas and values used to compute the association likelihoods $L_{\varphi}^{\mathbf{all}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j)$ in the case study presented in Section 3.

No rule If no prior geological knowledge is used, all the associations are assumed equally likely and

$$L_{\varphi_i}^{\mathbf{all}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = 1$$

for $i \in \{1, 2\}$.

Family rule only If only family rules are used, then $R_{\varphi}^{\mathbf{fam}}(\mathbb{v}_i) = 1$ and

$$L_{\varphi_i}^{\mathbf{all}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = R_{\varphi_i}^{\mathbf{fam}}(\mathbb{v}_i).$$

In this Santos Basin case study, the family rules rely on the dip orientation of the digitized fault evidence, and

$$R_{\varphi_1}^{\mathbf{fam}}(\mathbb{v}_i) = \begin{cases} 1, & \text{if the piece of fault evidence } \mathbb{v}_i \text{ is dipping toward the West, and} \\ 0, & \text{otherwise.} \end{cases}$$

As there is no uncertainty on which family the pieces of evidence belong to, the rule for the family φ_2 can be computed from the one for φ_1: $R_{\varphi_2}^{\mathbf{fam}}(\mathbb{v}_i) = 1 - R_{\varphi_1}^{\mathbf{fam}}(\mathbb{v}_i)$.

Orientation rules The orientation computed between two pieces of evidence is accounted using a rule and combined with the previously defined family rules using

$$L_{\varphi}^{\mathbf{all}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = R_{\varphi}^{\mathbf{fam}}(\mathbb{v}_i)\, R_{\varphi}^{\mathbf{fam}}(\mathbb{v}_j)\, R_{\varphi}^{\mathbf{assoc}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j),$$

with $R_{\varphi}^{\mathbf{assoc}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = R_{\varphi}^{orient}$ being a discrete association rule:

$$R_{\varphi_i}^{orient} = \begin{cases} 1, & \text{if the strike orientation between } \mathbb{v}_i \text{ and } \mathbb{v}_j \text{ is between } strike_{\varphi_i}^{min} \text{ and } strike_{\varphi_i}^{max}, \\ 0, & \text{otherwise,} \end{cases}$$

with $strike_{\varphi_1}^{min} = 330$, $strike_{\varphi_1}^{max} = 15$, $strike_{\varphi_2}^{min} = 145$, and $strike_{\varphi_2}^{max} = 195$.

All rules In this last case, a distance rule is also taken into account. The distance association likelihood is defined

$$R_{\varphi}^{dist}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = max(1 - dist(\mathbb{v}_i \leftrightarrow \mathbb{v}_j)/dist_{max}, 0),$$

with $dist_{max} = 3600$ being the dimension (in meter) of the longest fault observed in the area of interest. This association rule is combined with the orientation rule:

$$R_{\varphi}^{\mathbf{assoc}}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) = R_{\varphi}^{dist}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j) R_{\varphi}^{orient}(\mathbb{v}_i \leftrightarrow \mathbb{v}_j).$$

Rule falsifications In Section 3.4.2, the numerical values used for the orientation and distance rules are falsified to become:

$$\begin{cases} strike_{\varphi_1}^{min} = 330 + \alpha_{err}, \\ strike_{\varphi_1}^{max} = 15 + \alpha_{err}, \\ strike_{\varphi_2}^{min} = 145 + \alpha_{err}, \\ strike_{\varphi_2}^{max} = 195 + \alpha_{err}, \text{and} \\ dist_{max} = 3600 + \delta_{err}. \end{cases}$$

6.2 Sensitivity to scale parameters

The simulation process is parameterized by a scalar value α_{draw} and by a probability mass function P_{seg}, which both relate to the fault size and impact the number of simulated fault surfaces. Figure 9 exhibits statistics on the number of simulated faults, while modifying these two parameters. Statistics are computed over 5×10^5 realizations, with pieces of fault evidence extracted from the Santos Basin model, along 4 and 6 virtual cross-sections. All of the rules described in Section 3.2 are used for these experiments.

The probability mass function P_{seg} used to downscale a fault into several segments (Figure 3) clearly influences the mean number of simulated faults (Figure $9a_2, b_2$). As expected, the use of a linearly increasing density function leads to the simulation of more faults as compared to the decreasing function. When the number of fault observations increases, this trend becomes significant whereas the variability over the total set of realizations increases. For applications which seek to preferentially generate parsimonious scenarios in terms of number of faults, the choice of a decreasing mass function seems appropriate. In general, however, it seems more relevant to explore the search space, so we will choose the uniform mass function for P_{seg} from now on, as it almost spans the same extreme number of faults as with the increasing or decreasing mass functions.

The scalar α_{draw} defines the likelihood that major structures (containing more fault data) are selected earlier during the simulation (Equation 3 and step 3.2 in Figure 2). As expected, a negative or a low α_{draw} increases the number of generated fault structures. The mild decrease of this number when α_{draw} is larger than 1 suggests that the other factors tend to naturally limit the appearance of very large fault objects which would gather many fault evidences. This emerging behavior is qualitatively consistent with highly skewed distributions observed for fracture size distributions (Bonnet et al., 2001). In terms of range, choosing $\alpha_{draw} = 2$ approximately spans the same minimum and maximum number of faults as generated with other values, so we will keep a value of 2 in future experiments.

6.3 Outline of a hierarchical sampling strategy

We provide here the outline of a hierarchical sampling strategy, and hope it can serve as a basis for further research in stochastic structural interpretation:

1. Define the starting list of realisable associations $\mathbb{G}_0^{\mathbf{all}} \leftarrow \mathbb{G}^{\mathbf{all}}$ and the starting index $h \leftarrow 0$
2. Find the maximal cliques of $\mathbb{G}_h^{\mathbf{all}}$
3. Generate N_h scenarios by sampling a possible fault from the maximal cliques. Denote as $\mathbb{V}_{n_h}$, $n_h = 1, \ldots, N_h$ the set of graph vertices corresponding to that clique.
4. For each scenario $n_h = 1, \ldots, N_h$:
   (a) Increment the hierarchical level $h \leftarrow h + 1$
   (b) Update the current list of realisable associations: $\mathbb{G}_h^{\mathbf{all}} \leftarrow \mathbb{G}_{h-1}^{\mathbf{all}} \setminus \mathbb{V}_{n_h}$
   (c) If $\mathbb{G}_h^{\mathbf{all}}$ still has vertices, go to Step 2, otherwise terminate.

References

Álvarez-Gómez, J. A. (2019). Fmc—earthquake focal mechanisms data management,

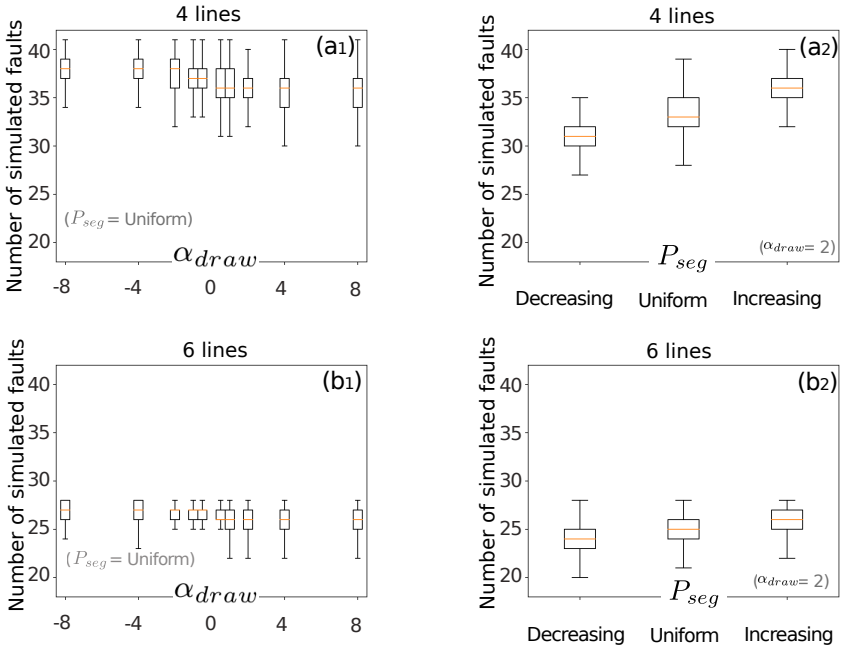

Figure 9. Sensitivity of the number of simulated faults to the parameter α_{draw} (a_1, b_1) and to the probability mass function (a_2, b_2) used while drawing the number of fault segments P_{seg}. Statistics computed over 5×10^5 realizations for the cases with 4 (a_1, a_2) and 6 (b_1, b_2) seismic lines.

771 cluster and classification. *SoftwareX*, *9*, 299–307.

772 Anquez, P., Pellerin, J., Irakarama, M., Cupillard, P., Lévy, B., & Caumon, G.
773 (2019, January). Automatic correction and simplification of geological maps
774 and cross-sections for numerical simulations. *Comptes Rendus Geoscience*,
775 *351*, 48-58. doi: 10.1016/j.crte.2018.12.001

776 Aydin, O., & Caers, J. K. (2017). Quantifying structural uncertainty on fault
777 networks using a marked point process within a Bayesian framework. *Tectono-*
778 *physics*, *712*, 101–124. doi: 10.1016/j.tecto.2017.04.027

779 Baudon, C., & Cartwright, J. A. (2008). 3D seismic characterisation of an array
780 of blind normal faults in the Levant Basin, Eastern Mediterranean. *Journal of*
781 *Structural Geology*, *30*(6), 746–760. doi: 10.1016/j.jsg.2007.12.008

782 Bond, C. (2015). Uncertainty in structural interpretation: Lessons to be learnt.
783 *Journal of Structural Geology*, *74*, 185–200. doi: 10.1016/j.jsg.2015.03.003

784 Bond, C., Gibbs, A., Shipton, Z., & Jones, S. (2007). What do you think this is?
785 "Conceptual uncertainty"in geoscience interpretation. *GSA Today*, *17*(11), 4.
786 doi: 10.1130/GSAT01711A.1

787 Bonneau, F., Caumon, G., & Renard, P. (2016). Impact of a stochastic sequential
788 initiation of fractures on the spatial correlations and connectivity of discrete
789 fracture networks. *Journal of Geophysical Research: Solid Earth*, *121*(8),
790 5641–5658. doi: 10.1002/2015JB012451

791 Bonnet, E., Bour, O., Odling, N. E., Davy, P., Main, I., Cowie, P., & Berkowitz, B.
792 (2001). Scaling of fracture systems in geological media. *Reviews of geophysics*,
793 *39*(3), 347-383.

794 Bron, C., & Kerbosch, J. (1973). Algorithm 457: finding all cliques of an undirected
795 graph. *Communications of the ACM*, *16*(9), 575–577. doi: 10.1145/362342
796 .362367

797 Caumon, G. (2010). Towards stochastic time-varying geological modeling. *Mathe-*
798 *matical Geosciences*, *42*(5), 555-569.

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*(366), 92–96.

Cherpeau, N., & Caumon, G. (2015). Stochastic structural modelling in sparse data situations. *Petroleum Geoscience*, *21*(4), 233–247. doi: 10.1144/petgeo2013-030

Cherpeau, N., Caumon, G., Caers, J., & Lévy, B. (2012). Method for stochastic inverse modeling of fault geometry and connectivity using flow data. *Mathematical Geosciences*, *44*(2), 147–168. doi: 10.1007/s11004-012-9389-2

Cherpeau, N., Caumon, G., & Lévy, B. (2010a). Stochastic simulation of fault networks from 2D seismic lines. In *SEG Expanded Abstracts* (Vol. 29, pp. 2366–2370). doi: 10.1190/1.3513325

Cherpeau, N., Caumon, G., & Lévy, B. (2010b). Stochastic simulations of fault networks in 3d structural modeling. *Comptes Rendus Géoscience*, *342*(9), 687–694. doi: 10.1016/j.crte.2010.04.008

Davy, P., Le Goc, R., & Darcel, C. (2013). A model of fracture nucleation, growth and arrest, and consequences for fracture density and scaling. *Journal of Geophysical Research: Solid Earth*, *118*(4), 1393–1407. doi: 10.1002/jgrb.50120

de la Varga, M., & Wellmann, J. F. (2016, August). Structural geologic modeling as an inference problem: A Bayesian perspective. *Interpretation*, *4*(3), SM1-SM16. doi: 10.1190/INT-2015-0188.1

Edwards, J., Lallier, F., Caumon, G., & Carpentier, C. (2018). Uncertainty management in stratigraphic well correlation and stratigraphic architectures: A training-based method. *Computers & Geosciences*, *111*, 1–17. doi: 10.1016/j.cageo.2017.10.008

Emerson. (2018). *SKUA-GOCAD.* https://www.pdgm.com/products/skua-gocad/.

Ferrill, D. A., Stamatakos, J. A., & Sims, D. (1999). Normal fault corrugation: Implications for growth and seismicity of active normal faults. *Journal of Struc-*

tural Geology, *21*(8), 1027–1038. doi: 10.1016/S0191-8141(99)00017-6

Freeman, B., Boult, P. J., Yielding, G., & Menpes, S. (2010). Using empirical geological rules to reduce structural uncertainty in seismic interpretation of faults. *Journal of Structural Geology*, *32*(11), 1668–1676. doi: 10.1016/j.jsg.2009.11.001

Freeman, B., Yielding, G., & Badley, M. (1990). Fault correlation during seismic interpretation. *First Break*, *8*(3), 87–95. doi: 10.3997/1365-2397.1990006

Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical science. *Geological Society of America Bulletin*, *107*(8), 960–968. doi: 10.1130/0016-7606(1995)107⟨0960:GRGAAI⟩2.3.CO;2

Gillespie, P., Walsh, J., & Watterson, J. (1992). Limitations of dimension and displacement data from single faults and the consequences for data analysis and interpretation. *Journal of Structural Geology*, *14*(10), 1157–1172. doi: 10.1016/0191-8141(92)90067-7

Giraud, J., Lindsay, M., Jessell, M., & Ogarko, V. (2019). Towards geologically reasonable lithological classification from integrated geophysical inverse modelling: Methodology and application case. *Solid Earth Discussions*, *2019*, 1–27. doi: 10.5194/se-2019-164

Godefroy, G., Caumon, G., Ford, M., Laurent, G., & Jackson, C. A.-L. (2017). A parametric fault displacement model to introduce kinematic control into modeling faults from sparse data. *Interpretation*, *6*(2), 1–48. doi: 10.1190/int-2017-0059.1

Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2019). Structural interpretation of sparse fault data using graph theory and geological rules. *Mathematical Geosciences*.

Gombert, B., Duputel, Z., Jolivet, R., Doubre, C., Rivera, L., & Simons, M. (2018, February). Revisiting the 1992 Landers earthquake: A Bayesian exploration of co-seismic slip and off-fault damage. *Geophysical Journal International*,

212(2), 839-852. doi: 10.1093/gji/ggx455

Gratier, J.-P., & Guillier, B. (1993, March). Compatibility constraints on folded and faulted strata and calculation of total displacement using computational restoration (UNFOLD program). *Journal of Structural Geology*, *15*(3-5), 391–402. doi: 10.1016/0191-8141(93)90135-W

Grose, L., Ailleres, L., Laurent, G., Armit, R., & Jessell, M. (2019). Inversion of geological knowledge for fold geometry. *Journal of Structural Geology*, *119*, 1–14. doi: 10.1016/j.jsg.2018.11.010

Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., & Cousin-Dechenaud, T. (2018). Inversion of structural geology data for fold geometry. *Journal of Geophysical Research: Solid Earth*, *123*(8), 6318–6333. doi: doi.org/10.1029/2017JB015177

Henza, A. A., Withjack, M. O., & Schlische, R. W. (2011). How do the properties of a pre-existing normal-fault population influence fault development during a subsequent phase of extension? *Journal of Structural Geology*, *33*(9), 1312–1324. doi: 10.1016/j.jsg.2011.06.010

Holden, L., Mostad, P., Nielsen, B. F., Gjerde, J., Townsend, C., & Ottesen, S. (2003). Stochastic structural modeling. *Mathematical Geology*, *35*(8), 899–914. doi: 10.1023/B:MATG.0000011584.51162.69

Holgate, N., Joldes, G. R., & Miller, K. (2015). Efficient visibility criterion for discontinuities discretised by triangular surface meshes. *Engineering analysis with boundary elements*, *58*, 1–6.

Hollund, K., Mostad, P., Nielsen, B. F., Holden, L., Gjerde, J., Contursi, M. G., ... Sverdrup, E. (2002). Havana - a fault modeling tool. *Norwegian Petroleum Society Special Publications*, *11*, 157–171.

Irakarama, M., Cupillard, P., Caumon, G., Sava, P., & Edwards, J. (2019). Appraising structural interpretations using seismic data—theoretical elements. *Geophysics*, *84*(2), N29–N40.

Irving, A., Chavanne, E., Faure, V., Buffet, P., & Barber, E. (2010). An uncertainty modelling workflow for structurally compartmentalized reservoirs. *Geological Society, London, Special Publications*, *347*(1), 283–299.

Jackson, C. A.-L., Jackson, M. P., & Hudec, M. R. (2015). Understanding the kinematics of salt-bearing passive margins: A critical test of competing hypotheses for the origin of the Albian Gap, Santos Basin, offshore Brazil. *Bulletin*, *127*(11-12), 1730–1751.

Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, J., Hillier, M., ... Martin, R. (2014). Next generation three-dimensional geologic modeling and inversion. *Economic Geology*, *18*, 261–272.

Julio, C. (2015). *Conditionnement de la modélisation stochastique 3D des réseaux de failles* (Unpublished doctoral dissertation). Universite de Lorraine.

Julio, C., Caumon, G., & Ford, M. (2015a). Impact of the en echelon fault connectivity on reservoir flow simulations. *Interpretation*, *3*(4), SAC23–SAC34. doi: 10.1190/INT-2015-0060.1

Julio, C., Caumon, G., & Ford, M. (2015b). Sampling the uncertainty associated with segmented normal fault interpretation using a stochastic downscaling method. *Tectonophysics*, *639*, 56–67. doi: 10.1016/j.tecto.2014.11.013

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Knuth, D. E. (2005). *The Art of Computer Programming, Volume 4: Generating all Combinations and Partitions, Fascicle 3.* Addison-Wesley Professional.

Lallier, F., Antoine, C., Charreau, J., Caumon, G., & Ruiu, J. (2013). Management of ambiguities in magnetostratigraphic correlation. *Earth and Planetary Science Letters*, *371*, 26–36. doi: 10.1016/j.epsl.2013.04.019

Laurent, G., Caumon, G., Bouziat, A., & Jessell, M. (2013, April). A parametric method to model 3D displacements around faults with volumetric vector fields. *Tectonophysics*, *590*, 83–93. doi: 10.1016/j.tecto.2013.01.015

Lecour, M., Cognot, R., Duvinage, I., Thore, P., & Dulac, J.-C. (2001). Mod-

elling of stochastic faults and fault networks in a structural uncertainty study. *Petroleum Geoscience*, *7*(S), S31–S42. doi: 10.1144/petgeo.7.S.S31

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., & Betts, P. G. (2012). Locating and quantifying geological uncertainty in three-dimensional models: Analysis of the Gippsland Basin, southeastern Australia. *Tectonophysics*, *546*, 10–27. doi: 10.1016/j.tecto.2012.04.007

Lines, L. R., & Newrick, R. T. (2004). *Fundamentals of Geophysical Interpretation.* Society of Exploration Geophysicists. doi: 10.1190/1.9781560801726

Mai, P. M., Galis, M., Thingbaijam, K. K. S., Vyas, J. C., & Dunham, E. M. (2017, September). Accounting for Fault Roughness in Pseudo-Dynamic Ground-Motion Simulations. *Pure and Applied Geophysics*, *174*(9), 3419-3450. doi: 10.1007/s00024-017-1536-8

Manighetti, I., Caulet, C., Barros, L., Perrin, C., Cappa, F., & Gaudemer, Y. (2015). Generic along-strike segmentation of Afar normal faults, East Africa: Implications on fault growth and stress heterogeneity on seismogenic fault planes. *Geochemistry, Geophysics, Geosystems*, *16*(2), 443–467. doi: 10.1002/2014GC005691

Mann, C. J. (1993). Uncertainty in geology. In J. C. Davis & U. C. E. Herzfeld (Eds.), *Computers in geology—25 years of progress* (pp. 241–254). Oxford University Press, Inc.

Manzocchi, T., Heath, A., Childs, C., Telles, I., & Carneiro, M. (2019). Modelling fault zone displacement partitioning for risking across-fault juxtaposition. In *81st eage conference and exhibition 2019.*

Meisling, K. E., Cobbold, P. R., & Mount, V. S. (2001). Segmentation of an obliquely rifted margin, campos and santos basins, southeastern brazil. *AAPG bulletin*, *85*(11), 1903–1924.

Munthe, K., Holden, L., Mostad, P., & Townsend, C. (1994). Modelling sub-seismic fault patterns using a Marked Point Process. In *Ecmor iv-4th european conference on the mathematics of oil recovery.* doi: 10.3997/2214-4609.201411151

Nicol, A., Walsh, J., Childs, C., & Manzocchi, T. (2020). The growth of faults. In *Understanding faults* (pp. 221–255). Elsevier.

Nixon, C. W., Sanderson, D. J., Dee, S. J., Bull, J. M., Humphreys, R. J., & Swanson, M. H. (2014). Fault interactions and reactivation within a normal-fault network at Milne Point, Alaska. *AAPG Bulletin*, *98*(10), 2081–2107. doi: 10.1306/04301413177

Osypov, K., Yang, Y., Fournier, A., Ivanova, N., Bachrach, R., Yarman, C. E., ... Woodward, M. (2013). Model-uncertainty quantification in seismic tomography: Method and applications. *Geophysical Prospecting*, *61*(6), 1114–1134.

Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., & Jessell, M. (2018, April). Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization. *Solid Earth*, *9*(2), 385–402. doi: 10.5194/se-9-385-2018

Pakyuz-Charrier, E., Jessell, M., Giraud, J., Lindsay, M., & Ogarko, V. (2019). Topological analysis in monte carlo simulation for uncertainty estimation. *Solid Earth Discussions*, *2019*, 1–37. doi: 10.5194/se-2019-78

Peacock, D., & Sanderson, D. (1991). Displacements, segment linkage and relay ramps in normal fault zones. *Journal of Structural Geology*, *13*(6), 721–733. doi: 10.1016/0191-8141(91)90033-F

Ragon, T., Sladen, A., & Simons, M. (2018). Accounting for uncertain fault geometry in earthquake source inversions–i: theory and simplified application. *Geophysical Journal International*, *214*(2), 1174–1190.

Ramsay, J. G., & Huber, M. I. (1987). *The techniques of modern structural geology: Folds and Fractures* (Vol. 2). Academic Press.

967 Richards, F. L., Richardson, N. J., Bond, C. E., & Cowgill, M. (2015). Interpre-
968 tational variability of structural traps: implications for exploration risk and
969 volume uncertainty. *Geological Society, London, Special Publications*, *421*(1),
970 7–27. doi: 10.1144/SP421.13

971 Riesner, M., Durand-Riard, P., Hubbard, J., Plesch, A., & Shaw, J. H. (2017, May).
972 Building Objective 3D Fault Representations in Active Tectonic Settings. *Seis-*
973 *mological Research Letters*, *88*(3), 831-839. doi: 10.1785/0220160192

974 Rivenæs, J. C., Otterlei, C., Zachariassen, E., Dart, C., & Sjøholm, J. (2005). A 3D
975 stochastic model integrating depth, fault and property uncertainty for plan-
976 ning robust wells, Njord Field, offshore Norway. *Petroleum Geoscience*, *11*(1),
977 57–65. doi: 10.1144/1354-079303-612

978 Røe, P., Georgsen, F., & Abrahamsen, P. (2014). An uncertainty model for fault
979 shape and location. *Mathematical Geosciences*, *46*(8), 957–969.

980 Rosenbaum, M. S., & Culshaw, M. G. (2003). Communicating the risks arising from
981 geohazards. *Journal of the Royal Statistical Society: Series A (Statistics in So-*
982 *ciety)*, *166*(2), 261–270. doi: 10.1111/1467-985X.00275

983 Rotevatn, A., Jackson, C. A.-L., Tvedt, A. B., Bell, R. E., & Blækkan, I. (2018, Au-
984 gust). How do normal faults grow? *Journal of Structural Geology*. doi: 10
985 .1016/j.jsg.2018.08.005

986 Sanfeliu, A., & Fu, K.-S. (1983). A distance measure between attributed relational
987 graphs for pattern recognition. *IEEE transactions on systems, man, and cyber-*
988 *netics*(3), 353–362.

989 Schaaf, A., & Bond, C. E. (2019). Quantification of uncertainty in 3-d seismic in-
990 terpretation: implications for deterministic and stochastic geomodelling and
991 machine learning. *Solid earth*.

992 Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, *1*(1), 27–64. doi:
993 10.1016/j.cosrev.2007.05.001

994 Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface sys-*

tems (Vol. 236). John Wiley & Sons.

Schneeberger, R., de La Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F., & Herwegh, M. (2017). Methods and uncertainty estimations of 3-d structural modelling in crystalline rocks: a case study. *Solid Earth*, *8*(5), 987. doi: 10.5194/se-8-987-2017

Seiler, A., Aanonsen, S. I., Evensen, G., & Rivenæs, J. C. (2010). Structural surface uncertainty modeling and updating using the ensemble Kalman filter. *SPE Journal*, *15*(04), 1–062. doi: 10.2118/125352-PA

Sepúlveda, I., Liu, P. L.-F., Grigoriu, M., & Pritchard, M. (2017, September). Tsunami hazard assessments with consideration of uncertain earthquake slip distribution and location: Tsumani Hazard and Uncertain Earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*(9), 7252-7271. doi: 10.1002/2017JB014430

Smith, T. F., & Waterman, M. S. (1980). New stratigraphic correlation techniques. *The Journal of Geology*, *88*(4), 451–457.

Stoyan, D., & Gloaguen, R. (2011, August). Nucleation and growth of geological faults. *Nonlinear Processes in Geophysics*, *18*(4), 529–536. doi: 10.5194/npg-18-529-2011

Suzuki, S., Caumon, G., & Caers, J. (2008). Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational Geosciences*, *12*(1), 105–119. doi: 10.1007/s10596-007-9063-9

Tal, Y., Hager, B. H., & Ampuero, J. P. (2018, January). The Effects of Fault Roughness on the Earthquake Nucleation Process. *Journal of Geophysical Research: Solid Earth*, *123*(1), 437-456. doi: 10.1002/2017JB014746

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, *2*(8), 492–494. doi: 10.1038/nphys375

Thiele, S. T., Jessell, M. W., Lindsay, M., Wellmann, J. F., & Pakyuz-Charrier, E.

(2016, October). The topology of geology 2: Topological uncertainty. *Journal of Structural Geology*, *91*, 74-87. doi: 10.1016/j.jsg.2016.08.010

Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T., & Cognot, R. (2002). Structural uncertainties: Determination, management, and applications. *Geophysics*, *67*(3), 840–852. doi: 10.1190/1.1484528

Torabi, A., & Berg, S. S. (2011). Scaling of fault attributes: A review. *Marine and Petroleum Geology*, *28*(8), 1444–1460. doi: 10.1016/j.marpetgeo.2011.04.003

Tvedt, A. B., Rotevatn, A., & Jackson, C. A. (2016). Supra-salt normal fault growth during the rise and fall of a diapir: Perspectives from 3D seismic reflection data, Norwegian North Sea. *Journal of Structural Geology*, *91*, 1–26. doi: 10.1016/j.jsg.2016.08.001

Walsh, J., Nicol, A., & Childs, C. (2002). An alternative model for the growth of faults. *Journal of Structural Geology*, *24*(11), 1669–1675. doi: 10.1016/S0191-8141(01)00165-1

Watterson, J. (1986). Fault dimensions, displacements and growth. *Pure and Applied Geophysics*, *124*(1-2), 365–373.

Wellmann, J. F., & Caumon, G. (2018, Nov). 3-D structural geological models: Concepts, methods, and uncertainties. In I. S. C. (Ed) (Ed.), *Advances in geophysics* (Vol. 59, p. 1-121). Elsevier. doi: 10.1016/bs.agph.2018.09.001

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

Wellmann, J. F., Lindsay, M., Poh, J., & Jessell, M. (2014). Validating 3-D structural models with geological knowledge for improved uncertainty evaluations. *Energy Procedia*, *59*, 374–381. doi: 10.1016/j.egypro.2014.10.391

Wilson, C. G., Bond, C. E., & Shipley, T. F. (2019). How can geologic decision making under uncertainty be improved? *Solid earth*, 1–34.

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., & Caumon, G. (2019, Octo-

ber). Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *GEOPHYSICS*, 1-48. doi: 10.1190/geo2019-0375.1

Yielding, G. (2016). The geometry of branch lines. *Geological Society, London, Special Publications*, *439*, SP439–1. doi: 10.1144/SP439.1

Zakian, P., Khaji, N., & Soltani, M. (2017, October). A Monte Carlo adapted finite element method for dislocation simulation of faults with uncertain geometry. *Journal of Earth System Science*, *126*(7), 105. doi: 10.1007/s12040-017-0878-z

Zehner, B., Börner, J. H., Görz, I., & Spitzer, K. (2015, June). Workflows for generating tetrahedral meshes for finite element simulations on complex geological structures. *Computers & Geosciences*, *79*, 105-117. doi: 10.1016/j.cageo.2015.02.009

Zhu, S., Hack, R., Turner, K., & Hale, M. (2003). How far will uncertainty of the subsurface limit the sustainability planning of the subsurface? In *Proc. sustainable development & management of the subsurface (sdms) conference* (pp. 5–7).