



HAL
open science

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

Gabriel Godefroy, Guillaume Caumon, Gautier Laurent, François Bonneau

► To cite this version:

Gabriel Godefroy, Guillaume Caumon, Gautier Laurent, François Bonneau. Multi-scenario interpretations from sparse fault evidence using graph theory and geological rules. *Journal of Geophysical Research: Solid Earth*, 2021, 126 (2), pp.e2020JB020022. 10.1029/2020JB020022 . hal-02562611v2

HAL Id: hal-02562611

<https://hal.univ-lorraine.fr/hal-02562611v2>

Submitted on 20 Feb 2021

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

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

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
¹F-54000 Nancy, France
²F-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.

53 Introduction

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

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

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

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

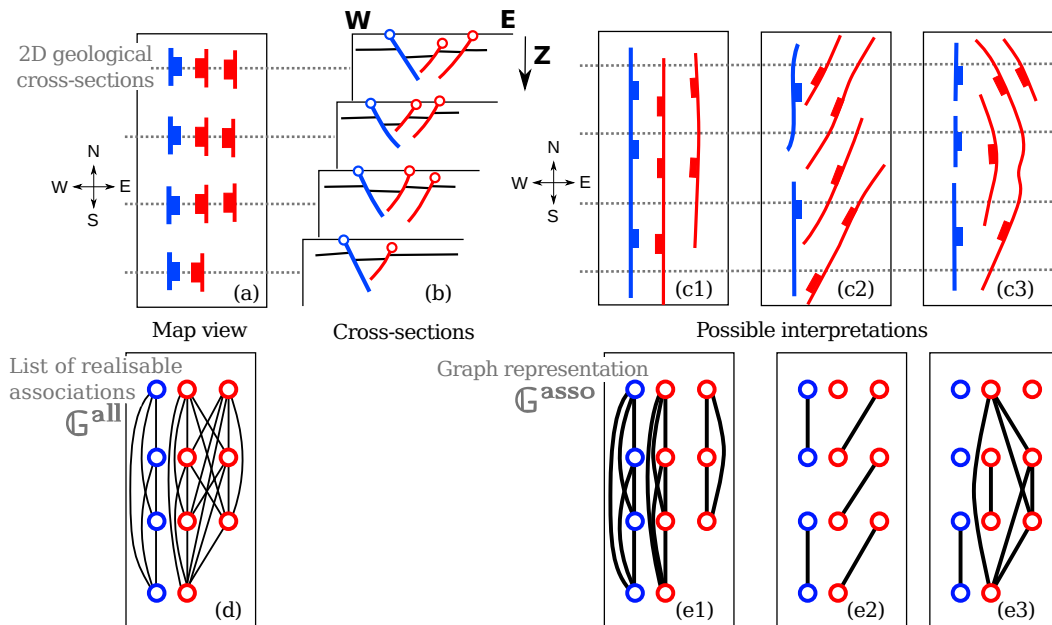


Figure 1. Associating labeled pieces of fault evidence (red: east-dipping and blue: west-dipping) interpreted (a) in map view or (b) on two-dimensional seismic lines is an under-constrained problem. (c₁, c₂ and c₃) Several structural interpretations are possible (Modified from Freeman et al., 1990; Godefroy et al., 2019). (d) In \mathbb{G}^{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. (e₁, e₂ and e₃) Plausible interpretations are represented by an association graph where the edges link pieces of fault evidence interpreted as belonging to the same fault.

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

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

120 **1 Structural uncertainty: state of the art**

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

126 To characterize structural uncertainty, one may ask a population of geologists to
127 interpret a particular data set (e.g., Bond et al., 2007; Schaaf & Bond, 2019). How-
128 ever, interpreting a subsurface data set in three dimensions commonly takes up to
129 several months, so this strategy is difficult to generalize. Alternatively, the promise
130 of 3D stochastic modeling is to use computing power to assess structural uncertainty.
131 Stochastic structural modeling has already been proposed to generate several scenar-
132 ios while taking account of seismic image quality and faults below seismic resolution
133 (Aydin & Caers, 2017; Hollund et al., 2002; Holden et al., 2003; Irving et al., 2010;
134 Julio et al., 2015a, 2015b; Lecour et al., 2001); uncertainty related to reflection seismic
135 acquisition and processing (Osypov et al., 2013; Thore et al., 2002); geological field
136 measurement uncertainty (Jessell et al., 2014; Lindsay et al., 2012; Pakyuz-Charrier et
137 al., 2019; Wellmann et al., 2014); structural parameters for folding (Grose et al., 2019,
138 2018); and observation gaps (Aydin & Caers, 2017; Cherpeau et al., 2010b; Cherpeau
139 & Caumon, 2015; Holden et al., 2003). Considering several structural interpretations
140 has also proved useful to propagate uncertainties to subsurface flow problems (Julio et
141 al., 2015b), to rank structural models against physical data and ultimately to falsify
142 some of the interpretations using a Bayesian approach (Cherpeau et al., 2012; de la
143 Varga & Wellmann, 2016; Irakarama et al., 2019; Seiler et al., 2010; Suzuki et al.,
144 2008; Wellmann et al., 2014).

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

156 ical convergence properties, but simulating fault networks in the presence of a large
 157 number of fault data remains computationally challenging. Therefore, Cherpeau et
 158 al. (2010b, 2012) and Cherpeau and Caumon (2015) propose a parsimonious method
 159 which anchors the first simulated faults to the available evidence, before simulating
 160 unseen fault objects. All these iterative stochastic fault models are difficult to use
 161 in practice, because of the combinatorial complexity of the problem (Godefroy et al.,
 162 2019; Julio, 2015), of the difficulty to integrate geological, kinematical and mechanical
 163 concepts into the stochastic model (Godefroy et al., 2017; Laurent et al., 2013; Nicol
 164 et al., 2020; Røe et al., 2014; Rotevatn et al., 2018), and of the geometric challenges to
 165 robustly build such three-dimensional structural models (e.g., due to meshing issues,
 166 see Anquez et al., 2019; Zehner et al., 2015).

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

171 **2 Multi-scenario interpretations using graph decomposition**

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

188 **2.1 Accounting for geological knowledge**

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

192 **2.1.1 Fault families**

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

201 faults active at different geological times. Interactions with other faults (Henza et al.,
 202 2011), syn-sedimentary structures or salt diapirs (Tvedt et al., 2016) also guide struc-
 203 tural interpretation. Data obtained by inversion of focal mechanisms (Álvarez-Gómez,
 204 2019) and mineralization observed within the fault core during mapping can also be
 205 used to cluster the pieces of evidence into distinct families.

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

$$\mathbb{G}^{\text{all}} = \bigcup_{\varphi=1}^{\varphi=n} \mathbb{G}_{\varphi}^{\text{all}}, \quad (1)$$

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

221 *2.1.2 Associating pieces of evidence*

222 To determine which pieces of evidence belong to the same fault, geological concepts
 223 including expected fault orientation, size, kinematic or corrugation are used. The
 224 broad literature on scaling laws (Bonnet et al., 2001; Torabi & Berg, 2011) can also
 225 help determining whether two observations separated by a given distance may belong
 226 to the same fault. Kinematic analysis methods, such as the analysis of the throw
 227 distribution along strike, are also used during structural interpretation (Nixon et al.,
 228 2014; Tvedt et al., 2016). Well or seismic data on which no fault can be interpreted
 229 may also help interpreting that two pieces of evidence cannot belong to the same fault.

230 To include such a geological knowledge in the sampler, an association rule $R_{\varphi}^{\text{assoc}}(\mathbf{v}_i \leftrightarrow$
 231 $\mathbf{v}_j)$ quantifies the likelihood that two pieces of evidence (\mathbf{v}_i and \mathbf{v}_j) of the same family
 232 (φ) belong to the same fault (Godefroy et al., 2019). Association rules are defined
 233 from general structural concepts about faults and from some geometric characteristics
 234 associated with each fault family. An association rule $R_{\varphi}^{\text{assoc}}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j)$ returns a num-
 235 ber between 0 (\mathbf{v}_i and \mathbf{v}_j cannot belong to the same fault of the family φ) and 1 (if
 236 both fault data are likely to belong to the same fault).

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

246 **2.2 Size of the search space**

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

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

260 **2.3 Sampling by stochastic graph partitioning**

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

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

275 **Creation and segmentation of the set of realisable association:** the family and
 276 association rules are used to compute a graph of all realisable associations $\mathbb{G}_\varphi^{\text{all}}$
 277 for each family φ (Figure 2, step 2.1). A wide range of formulas can be used to
 278 compute the likelihoods according to the defined rules. Each edge $e^\varphi(\mathbf{v}_i, \mathbf{v}_j)$ of
 279 $\mathbb{G}_\varphi^{\text{all}}$ (linking two pieces of evidence \mathbf{v}_i and \mathbf{v}_j) carries an association likelihood
 280 $L_\varphi^{\text{all}}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j)$ for each family φ :

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

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

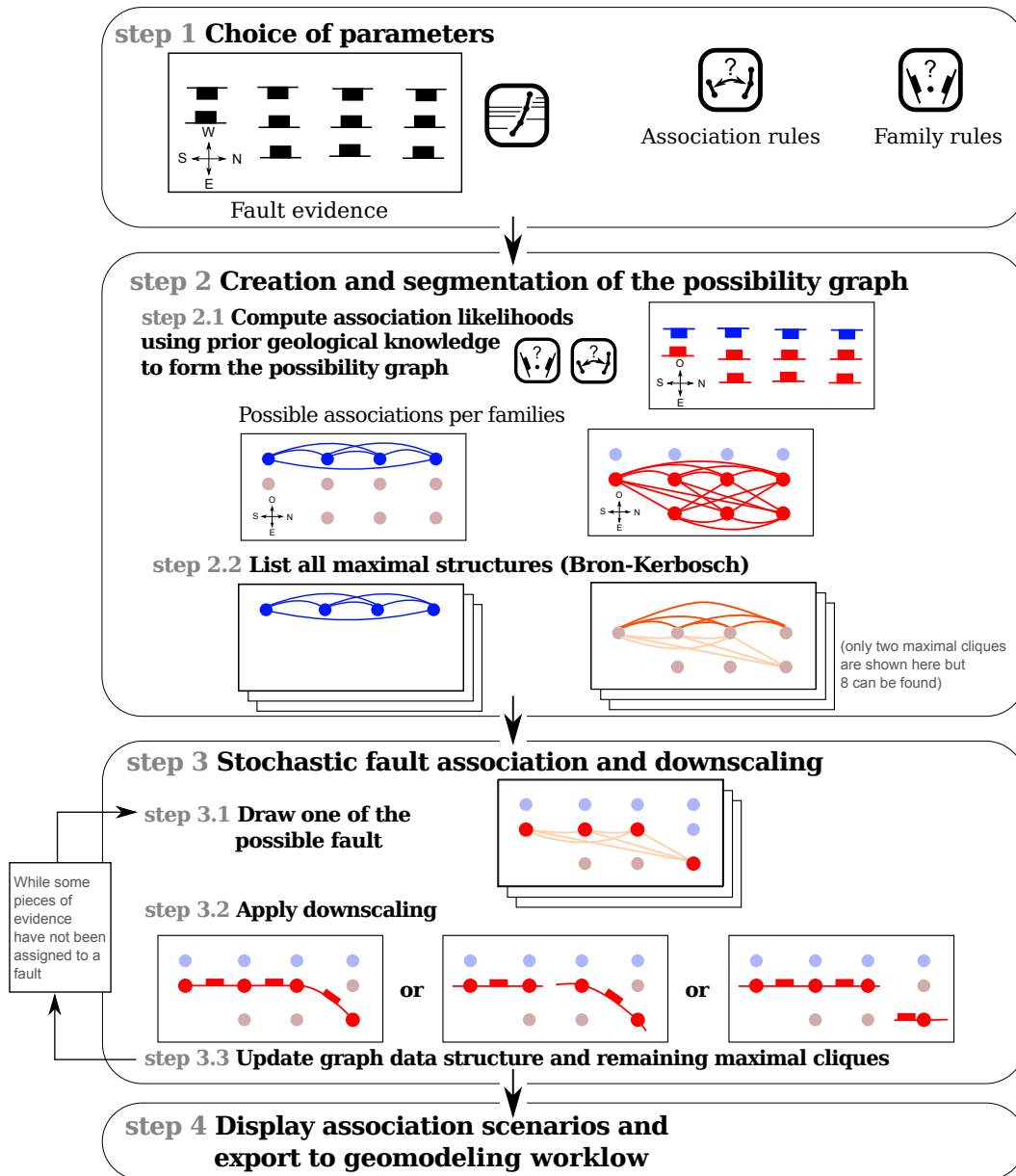


Figure 2. Sequential stochastic algorithm interpreting the available fault data as distinct cliques in $\mathbb{G}_\varphi^{\text{all}}$. (step 1) The algorithm requires input fault interpretations and a set of interpretation rules. (step 2.1) Family and association rules are used to compute the graphs $\mathbb{G}_\varphi^{\text{all}}$ 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 $\mathbb{G}_\varphi^{\text{all}}$ 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.

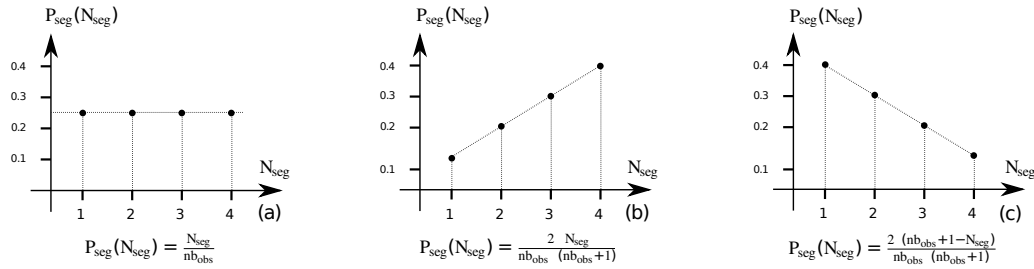


Figure 3. Three probability mass functions that can be used to draw the number of segments during the segmentation step: (a) uniform, (b) increasing, and (c) decreasing.

293 **Stochastic fault association:** cliques are randomly and sequentially drawn and re-
 294 moved from the remaining realisable associations until each fault evidence has
 295 been assigned to a fault. Several strategies can be defined and chosen for this
 296 sequential random selection of faults. To mimic the interpretation process by
 297 experts, who tend to first focus on the major structures (e.g., Lines & Newrick,
 298 2004), we propose to preferentially select large and overall likely faults before
 299 selecting small and unlikely faults. At each step, the sampling probability of
 300 a clique $\mathbb{F} = \{v_i, \dots, v_j\}$ depends on the number of nodes $|\mathbb{F}|$ and on the mean
 301 association likelihood $\overline{L_\varphi^{\text{all}}(\mathbb{F})}$ as

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

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

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

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

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

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

336 When fault data are interpreted along parallel two-dimensional seismic lines, an
337 extra constraint can be added to prevent fault-fault intersections. The obtained associ-
338 ations are then free of branch lines which. In the sequential fault association framework,
339 after each downscaling step (step 3.2), the edges crossing the simulated fault are re-
340 moved from the remaining cliques while updating the data structure (step 3.3). A sim-
341 ilar graph problem is solved using the Dynamic Time Warping algorithm (Levenshtein,
342 1966) in order to correlate stratigraphic data along wells (Edwards et al., 2018; Lallier
343 et al., 2013; Smith & Waterman, 1980). Note that the proposed strategy to avoid
344 intersections is applicable only in the presence of fault interpretations made on par-
345 allel sections. Extending this constraint to irregularly sampled sparse data could be
346 achieved, for example, by using visibility criteria (Holgate et al., 2015).

347 **3 Application to sparse data from Santos basin, offshore Brazil**

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

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

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

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

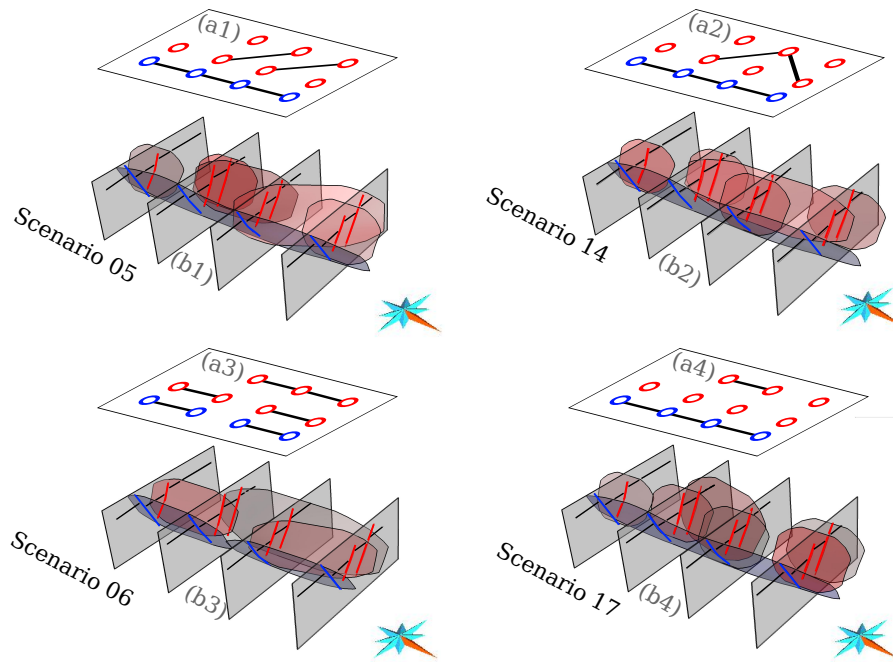


Figure 4. (a) The generated interpretation scenarios are represented by association graphs.

(b) For each clique, a fault surface can be interpolated using the geometry of the fault ev-

idence. Data and generated models can be visualized at: [gabrielgodefroy.github.io/](https://github.com/gabrielgodefroy/StochasticInterpData/Fig4/Fig4.html)

[StochasticInterpData/Fig4/Fig4.html](https://github.com/gabrielgodefroy/StochasticInterpData/Fig4/Fig4.html)

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

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

393 3.1 Synthetic two-dimensional lines

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

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

407 3.2 Association rules

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

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

and

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

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

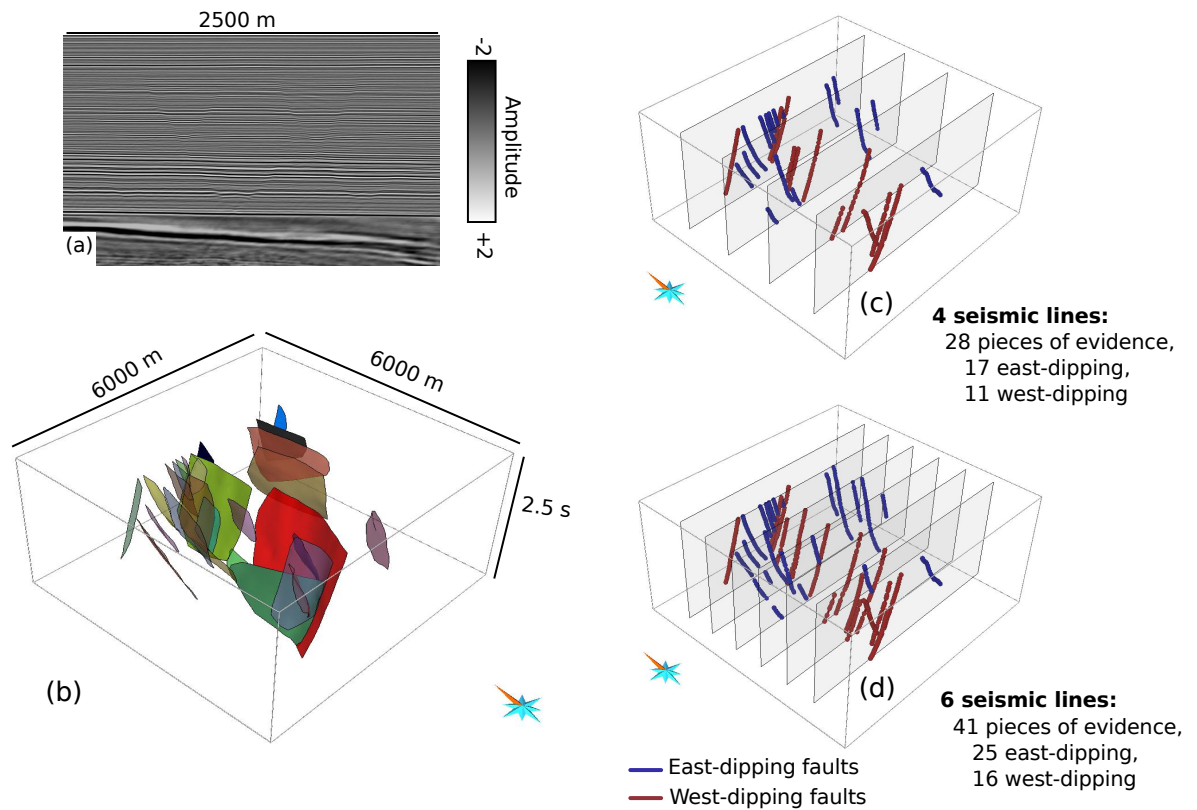


Figure 5. Reference structural model located in the Santos Basin, offshore Brazil. (a) Available reflection seismic data (courtesy of PGS). (b) Reference fault network. (c, d) Generated interpreted synthetic parallel two-dimensional seismic lines. An interactive three-dimensional model is provided as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

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

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

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

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

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

421 Finally, both rules are combined using:

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

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

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

434 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}_l^{asso} and \mathbb{G}_m^{asso} . N_{diff} is a special
 case of graph edit distance (Sanfeliu & Fu, 1983), in which the only edit operations
 are edge insertion and deletion:

$$N_{diff}(l, m) = \sum_{e \in \text{edges}} d_{l,m}(e) \quad (4)$$

where

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

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

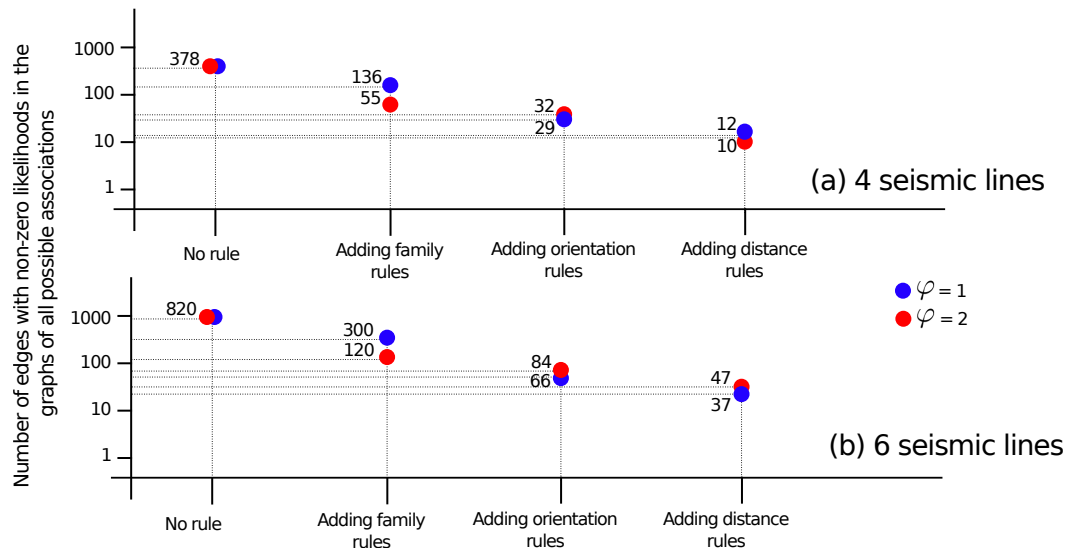


Figure 6. Number of edges per fault family (φ) in the graphs of all possible associations ($\mathbb{G}_\varphi^{\text{all}}$) for evidence extracted along (a) 4 seismic lines, and (b) 6 seismic lines. The integration of geological rules (Appendix 6.1) reduces the density of the graphs of all possible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

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

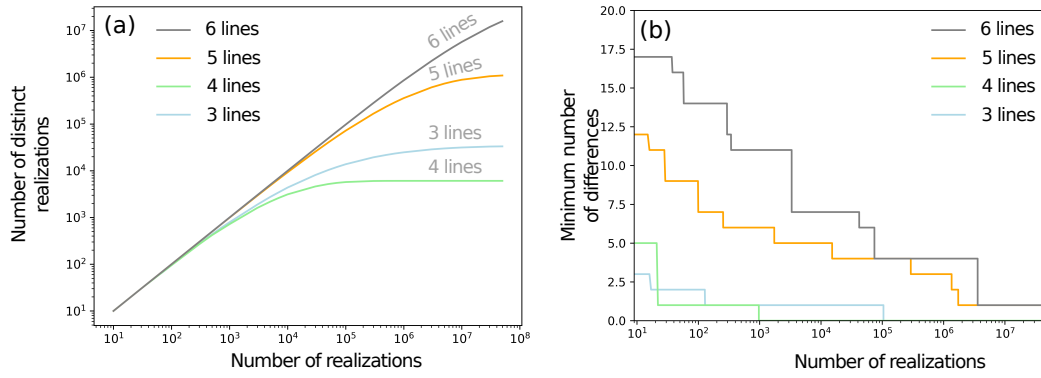


Figure 7. (a) Number of distinct association scenarios obtained from 5×10^7 realizations from 3, 4, 5, and 6 seismic lines. A plateau is not reached for the case with 5 and 6 seismic lines, highlighting the very large computational complexity of the problem. (b) Minimum number of differences to the reference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

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

441 For the case with 3 seismic lines, the plateau is not yet fully reached but for
 442 the case with 4 seismic lines, a plateau of 6072 distinct realizations is reached after
 443 3×10^6 realizations. In both cases, the reference association is found 2026 times and
 444 3164 times for 3 and 4 seismic lines, respectively. This can be explained by the differ-
 445 ent information content carried by these data sets (see Figure 6 of the supplementary
 446 material). When simulating from 5 and 6 seismic lines (35 and 41 fault data, respec-
 447 tively), the numbers of distinct scenarios is still significantly increasing after 5×10^7
 448 realizations and the reference association is not found. In both cases, the best scenario
 449 produced has a number of differences N_{diff} equal to 1 (Figure 7.b).

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

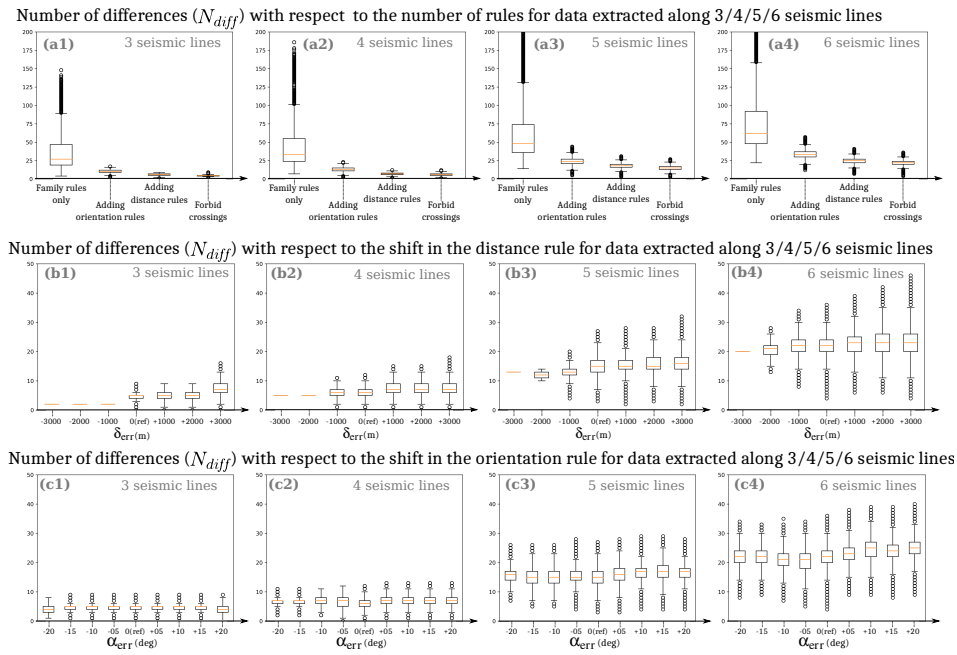


Figure 8. Box plots showing the minimum, mean and maximum number of differences (N_{diff}) between the simulated association graphs and the reference one. Graphs computed for sparse data extracted from 3 ($a_1 - c_1$) to 6 ($a_4 - c_4$) seismic lines. ($a_1 - a_4$) The integration of more geological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

462 The experiment shows that the formalism and the chosen metric are favorable
 463 when relatively few pieces of fault evidence are available. This may seem counter-
 464 intuitive; however, when pieces of evidence are added, the difficulty to explore a larger
 465 search space increases in a non-polynomial way, as also observed by Edwards et al.
 466 (2018) for well correlation.

467 3.4 Influence of the chosen geological rules on simulation results

468 3.4.1 Impact of the number of rules

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

473 To analyze these results, we now consider the number of differences $N_{diff}(l, ref)$
 474 between each simulated association \mathbb{G}^{asso^l} and the reference association \mathbb{G}_{ref}^{asso} inter-
 475 preted from the full 3D seismic data set. When the simulations are run with more
 476 rules (Figure 8.a₁ – a₄), the minimum, mean, and maximum number of differences
 477 $N_{diff}(l, ref)$ consistently decrease. Moreover, we observe that rules interact with the
 478 data density on two respects. First, considering a single rule in sparse data settings
 479 (cases with 3 or 4 sections) yields realizations closer to the reference than considering
 480 more informative rules in denser data settings (cases with 5 or more sections). This
 481 can be explained by the non-convergence of the sampler for 5 and 6 sections, as shown
 482 in Figure 7.a. Second, for a given set of rules, the absolute number of differences from
 483 the reference increases with the number of data. This is a direct effect of the increasing
 484 complexity of the search space, and may also be explained by the non-convergence of
 485 the sampler with 5×10^7 realizations for more than 5 seismic lines (Figure 8.a₄).

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

491 3.4.2 *Deliberately selecting biased rules*

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

498 We study the mean number of differences to the reference association (over 5×10^5
 499 realizations) according to these inappropriate choices (Figure 8.b,c). Strongly biased
 500 rules make it impossible to retrieve the reference association from fault evidence sam-
 501 pled along 3, 4, 5 or 6 seismic lines (Figure 8.b₁, b₂, c₁, c₂, b₃, b₄ and c₃, c₄, respectively).
 502 When simulating interpretation scenarios from data extracted along 5 or 6 seismic sec-
 503 tions, the reference scenario is never retrieved.

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

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

513 4 Discussion and ways forward

514 4.1 On uncertainty and objectivity

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

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

530 A major difference between the proposed approach and the expert-based manual
531 interpretation is the intrinsic ability of the former to generate scenarios, whereas most
532 of the latter tends to end up with deterministic solution. Geoscience is a discipline
533 where uncertainty is a key component (Frodeman, 1995) and geoscientists and inter-
534 preters are exposed to uncertainty from their first university courses. However, there
535 is a large loss between how geologists see uncertainty and what is often produced as
536 a result of their work. This is shown even in experiments assessing the interpretation
537 uncertainty, which ask a set of geoscientists to produce one interpretation each (Bond
538 et al., 2007; Bond, 2015; Schaaf & Bond, 2019). In most studies, a single determinis-
539 tic structural model is produced, and structural uncertainty are communicated in the
540 form of reports and diagrams. Cognitive biases also explain the difficulty of one to
541 work with a large number of hypotheses in parallel (Chamberlin, 1890; Wilson et al.,
542 2019). Cognitive biases have been demonstrated experimentally (Kahneman, 2011)
543 and are a possible explanation for interpretation bias (Bond et al., 2007). The advent
544 of computer-based methods makes it easier to explore aleatory uncertainties by per-
545 turbing a reference model (see Wellmann & Caumon, 2018, and references therein),
546 but addressing epistemic uncertainties is more challenging as it requires to formalize
547 the geological concepts. The graph-based sampling method proposed in this paper
548 clearly belongs to this latter class of methods.

549 A second major difference with classical fault interpretation is that our approach
550 relies on the interpreter to explicitly formulate and organize elementary association
551 rules, whereas expert based knowledge is seldom explicitly described. The order by
552 which we processed and combined the rules is driven in part by mathematical and al-
553 gorithmic convenience, and we do not claim it to be fully objective. Indeed, geological
554 interpretation always depends on the current state of knowledge and experience of an
555 interpreter, on some model assumptions, and on the tools used during the interpre-
556 tation (Bond et al., 2007; Chamberlin, 1890; Frodeman, 1995; Wellmann & Caumon,
557 2018). Nonetheless, as compared to classical expert-based interpretation, we see the
558 general approach proposed in this paper as a step towards making the interpreta-
559 tion process more transparent and reproducible by expressing formally the geological
560 concepts.

561 **4.2 On graphs for structural uncertainty assessment**

562 We see the graph-based method proposed in this paper as a possible way to com-
563 plement or generalize structural uncertainty assessment method. It starts from existing
564 observations and thus follows the same philosophy as perturbation strategies, which
565 consider uncertainties in the location and/or orientation of observations (Lindsay et
566 al., 2012; Pakyuz-Charrier et al., 2018; Wellmann et al., 2010) or modeled structures
567 (Holden et al., 2003; Lecour et al., 2001; Røe et al., 2014). Our approach starts with-
568 out any particular assumption about how to associate incomplete fault observations

569 together. Unlike previous iterative methods (Aydin & Caers, 2017; Cherpeau et al.,
570 2010a; Cherpeau & Caumon, 2015), the potential major fault structures are processed
571 in the early steps of our algorithm thanks to the maximal clique detection, assuming
572 that largest faults are most likely to correspond to many graph nodes. This reproduces
573 a classical interpretation process whereby geologists focus on largest structures before
574 focusing on smaller objects (e.g., Lines & Newrick, 2004).

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

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

600 **4.3 Are the produced interpretations “geologically realistic”?**

601 Another major difference between expert-based interpretations and the proposed
602 method concerns the amount of interpretive concepts used. Interpretation concepts
603 typically arise from outcrop or subsurface analog data bases, laboratory and numerical
604 models, or on an inference process applied directly to the data at hand. Those con-
605 cepts are conveyed in the form of text, sketches and oral presentations. The proposed
606 graph-based method calls for formally defining the rules according to each specific
607 geological context. In this paper, we only tested relatively simple geological rules in
608 order to assess the consistency of the sampling algorithm. If 3-D models were built,
609 we could test also their validity by asking a population of geologist to visually inspect
610 the models and rate their likelihood. Further studies are clearly needed to help geosci-
611 entists defining the numerical interpretation rules corresponding to the interpretation
612 concepts. In the case of faulted structures, concepts include choosing fault surface ori-
613 entations from analog data sets (Aydin & Caers, 2017), accounting for fault curvature
614 and lateral extension, estimating the fault slip (Cherpeau & Caumon, 2015; Røe et
615 al., 2014), and evaluating fault segmentation (Julio et al., 2015a; Manighetti et al.,
616 2015; Manzocchi et al., 2019). Machine learning could also come into play in this pro-
617 cess, either by inferring rule parameters or a posteriori assessing the likelihood of the
618 various realizations produced by the sampling method. Training for these approaches

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

621 To test the proposed sampler, we started by choosing an ideal case where the
622 parameters of simple rules were calibrated directly on a reference model. We acknowl-
623 edge that this is never the case in practice where no reference model exists, but this
624 allowed us to compare the simulated scenario with a reference. Tests made in Section
625 3.4.2 suggest that, when rule parameters are chosen inappropriately, the simulated
626 models can show significant bias, and we can expect that this observation would also
627 hold for more complex rules. However, such an automatic method has no guarantee to
628 produce the same results as several interpretations made by several experts. One rea-
629 son is that the sampling method is likely to miss some important aspects of geological
630 interpretation. On this regard, we see two main avenues for improvement:

- 631 • First, the methodology does not completely automate the three-dimensional
632 structural modeling. Branch lines are not handled and a boundary represen-
633 tation of the subsurface is not generated. This makes it difficult to assess the
634 likelihood of the generated fault networks using advanced structural analyzes
635 such as global displacement analysis (Freeman et al., 2010) or structural restora-
636 tion (Gratier & Guillier, 1993). For this, fault geometries should be modeled
637 using explicit surfaces (Lecour et al., 2001; Røe et al., 2014) or implicit surfaces
638 (Aydin & Caers, 2017; Cherpeau et al., 2010b). Then, the geological formations
639 affected by the fault network should be modeled. Generating such geometries
640 would also be useful to assess the impact of fault network uncertainty on resource
641 assessment (Richards et al., 2015), to incorporate this source of uncertainty in
642 geophysical inverse problems (Giraud et al., 2019; Ragon et al., 2018), or to con-
643 sider the geological likelihood in a Bayesian inference problem (Caumon, 2010;
644 de la Varga & Wellmann, 2016).
- 645 • Second, the graph formalism at this stage only considers pairwise associations
646 but does not use the likelihood of associating several pieces of evidence at once.
647 This calls for translating geological concepts into numerical rules which apply
648 to all fault evidence at once. For example, one could consider the throw dis-
649 tribution along fault strike (as in Cherpeau & Caumon, 2015; Freeman et al.,
650 1990) or statistical relationships between observed separations and fault size
651 (e.g., Gillespie et al., 1992; Torabi & Berg, 2011). The association likelihoods
652 should also be updated during the graph-based interpretation to account for the
653 expected likelihood of fault-fault interactions.

654 **4.4 Inverse problem and clustering of structural interpretations**

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

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

673 5 Conclusions

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

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

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

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

703 Acknowledgments

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

713 **6 Appendices**714 **6.1 Rules expression**

715 We give here the numerical formulas and values used to compute the association
716 likelihoods $L_{\varphi}^{\text{all}}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j)$ 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

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

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

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

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

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

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

720 As there is no uncertainty on which family the pieces of evidence belong to,
721 the rule for the family φ_2 can be computed from the one for φ_1 : $R_{\varphi_2}^{\text{fam}}(\mathbf{v}_i) =$
722 $1 - R_{\varphi_1}^{\text{fam}}(\mathbf{v}_i)$.

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

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

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

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

724 with $\textit{strike}_{\varphi_1}^{\text{min}} = 330$, $\textit{strike}_{\varphi_1}^{\text{max}} = 15$, $\textit{strike}_{\varphi_2}^{\text{min}} = 145$, and $\textit{strike}_{\varphi_2}^{\text{max}} = 195$.

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

$$R_{\varphi}^{\text{dist}}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j) = \max(1 - \textit{dist}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j) / \textit{dist}_{\text{max}}, 0),$$

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

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

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

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

727

6.2 Sensitivity to scale parameters

728

729

730

731

732

733

734

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.

735

736

737

738

739

740

741

742

743

744

745

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 9a₂, b₂). 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.

746

747

748

749

750

751

752

753

754

755

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.

756

6.3 Outline of a hierarchical sampling strategy

757

758

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:

759

760

761

762

763

764

765

766

767

768

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

769

References

770

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

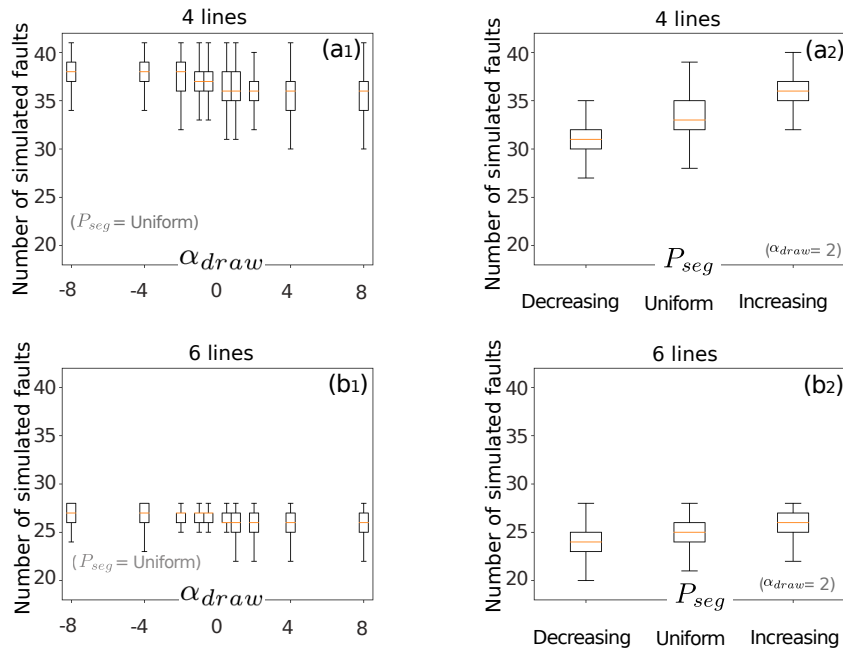


Figure 9. Sensitivity of the number of simulated faults to the parameter α_{draw} (a_1, b_1) and to the probability mass function (a_2, b_2) used while drawing the number of fault segments P_{seg} . Statistics computed over 5×10^5 realizations for the cases with 4 (a_1, a_2) and 6 (b_1, b_2) seismic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

- 799 Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*,
800 15(366), 92–96.
- 801 Cherpeau, N., & Caumon, G. (2015). Stochastic structural modelling in sparse data
802 situations. *Petroleum Geoscience*, 21(4), 233–247. doi: 10.1144/petgeo2013
803 -030
- 804 Cherpeau, N., Caumon, G., Caers, J., & Lévy, B. (2012). Method for stochastic in-
805 verse modeling of fault geometry and connectivity using flow data. *Mathemati-
806 cal Geosciences*, 44(2), 147–168. doi: 10.1007/s11004-012-9389-2
- 807 Cherpeau, N., Caumon, G., & Lévy, B. (2010a). Stochastic simulation of fault net-
808 works from 2D seismic lines. In *SEG Expanded Abstracts* (Vol. 29, pp. 2366–
809 2370). doi: 10.1190/1.3513325
- 810 Cherpeau, N., Caumon, G., & Lévy, B. (2010b). Stochastic simulations of fault
811 networks in 3d structural modeling. *Comptes Rendus Géoscience*, 342(9), 687–
812 694. doi: 10.1016/j.crte.2010.04.008
- 813 Davy, P., Le Goc, R., & Darcel, C. (2013). A model of fracture nucleation, growth
814 and arrest, and consequences for fracture density and scaling. *Journal of Geo-
815 physical Research: Solid Earth*, 118(4), 1393–1407. doi: 10.1002/jgrb.50120
- 816 de la Varga, M., & Wellmann, J. F. (2016, August). Structural geologic modeling
817 as an inference problem: A Bayesian perspective. *Interpretation*, 4(3), SM1-
818 SM16. doi: 10.1190/INT-2015-0188.1
- 819 Edwards, J., Lallier, F., Caumon, G., & Carpentier, C. (2018). Uncertainty
820 management in stratigraphic well correlation and stratigraphic architec-
821 tures: A training-based method. *Computers & Geosciences*, 111, 1–17. doi:
822 10.1016/j.cageo.2017.10.008
- 823 Emerson. (2018). *SKUA-GOCAD*. [https://www.pdgm.com/products/skua-
824 -gocad/](https://www.pdgm.com/products/skua-gocad/).
- 825 Ferrill, D. A., Stamatakos, J. A., & Sims, D. (1999). Normal fault corrugation: Im-
826 plications for growth and seismicity of active normal faults. *Journal of Struc-*

- 827 *tural Geology*, 21(8), 1027–1038. doi: 10.1016/S0191-8141(99)00017-6
- 828 Freeman, B., Boulton, P. J., Yielding, G., & Menpes, S. (2010). Using empiri-
829 cal geological rules to reduce structural uncertainty in seismic interpre-
830 tation of faults. *Journal of Structural Geology*, 32(11), 1668–1676. doi:
831 10.1016/j.jsg.2009.11.001
- 832 Freeman, B., Yielding, G., & Badley, M. (1990). Fault correlation during seismic in-
833 terpretation. *First Break*, 8(3), 87–95. doi: 10.3997/1365-2397.1990006
- 834 Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical
835 science. *Geological Society of America Bulletin*, 107(8), 960–968. doi: 10.1130/
836 0016-7606(1995)107<0960:GRGAAI>2.3.CO;2
- 837 Gillespie, P., Walsh, J., & Watterson, J. (1992). Limitations of dimension and
838 displacement data from single faults and the consequences for data analysis
839 and interpretation. *Journal of Structural Geology*, 14(10), 1157–1172. doi:
840 10.1016/0191-8141(92)90067-7
- 841 Giraud, J., Lindsay, M., Jessell, M., & Ogarko, V. (2019). Towards geologically rea-
842 sonable lithological classification from integrated geophysical inverse modelling:
843 Methodology and application case. *Solid Earth Discussions*, 2019, 1–27. doi:
844 10.5194/se-2019-164
- 845 Godefroy, G., Caumon, G., Ford, M., Laurent, G., & Jackson, C. A.-L. (2017).
846 A parametric fault displacement model to introduce kinematic control
847 into modeling faults from sparse data. *Interpretation*, 6(2), 1–48. doi:
848 10.1190/int-2017-0059.1
- 849 Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2019). Structural interpre-
850 tation of sparse fault data using graph theory and geological rules. *Mathemati-
851 cal Geosciences*.
- 852 Gombert, B., Duputel, Z., Jolivet, R., Doubre, C., Rivera, L., & Simons, M. (2018,
853 February). Revisiting the 1992 Landers earthquake: A Bayesian exploration
854 of co-seismic slip and off-fault damage. *Geophysical Journal International*,

- 855 212(2), 839-852. doi: 10.1093/gji/ggx455
- 856 Gratier, J.-P., & Guillier, B. (1993, March). Compatibility constraints on folded
857 and faulted strata and calculation of total displacement using computational
858 restoration (UNFOLD program). *Journal of Structural Geology*, 15(3-5),
859 391–402. doi: 10.1016/0191-8141(93)90135-W
- 860 Grose, L., Ailleres, L., Laurent, G., Armit, R., & Jessell, M. (2019). Inversion of ge-
861 ological knowledge for fold geometry. *Journal of Structural Geology*, 119, 1–14.
862 doi: 10.1016/j.jsg.2018.11.010
- 863 Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., & Cousin-Dechenaud,
864 T. (2018). Inversion of structural geology data for fold geometry. *Journal of*
865 *Geophysical Research: Solid Earth*, 123(8), 6318–6333. doi: doi.org/10.1029/
866 2017JB015177
- 867 Henza, A. A., Withjack, M. O., & Schlische, R. W. (2011). How do the properties
868 of a pre-existing normal-fault population influence fault development dur-
869 ing a subsequent phase of extension? *Journal of Structural Geology*, 33(9),
870 1312–1324. doi: 10.1016/j.jsg.2011.06.010
- 871 Holden, L., Mostad, P., Nielsen, B. F., Gjerde, J., Townsend, C., & Ottesen, S.
872 (2003). Stochastic structural modeling. *Mathematical Geology*, 35(8), 899–914.
873 doi: 10.1023/B:MATG.0000011584.51162.69
- 874 Holgate, N., Joldes, G. R., & Miller, K. (2015). Efficient visibility criterion for dis-
875 continuities discretised by triangular surface meshes. *Engineering analysis with*
876 *boundary elements*, 58, 1–6.
- 877 Hollund, K., Mostad, P., Nielsen, B. F., Holden, L., Gjerde, J., Contursi, M. G., ...
878 Sverdrup, E. (2002). Havana - a fault modeling tool. *Norwegian Petroleum*
879 *Society Special Publications*, 11, 157–171.
- 880 Irakarama, M., Cupillard, P., Caumon, G., Sava, P., & Edwards, J. (2019). Apprais-
881 ing structural interpretations using seismic data—theoretical elements. *Geo-*
882 *physics*, 84(2), N29–N40.

- 883 Irving, A., Chavanne, E., Faure, V., Buffet, P., & Barber, E. (2010). An uncertainty
884 modelling workflow for structurally compartmentalized reservoirs. *Geological*
885 *Society, London, Special Publications*, 347(1), 283–299.
- 886 Jackson, C. A.-L., Jackson, M. P., & Hudec, M. R. (2015). Understanding the kine-
887 matics of salt-bearing passive margins: A critical test of competing hypotheses
888 for the origin of the Albian Gap, Santos Basin, offshore Brazil. *Bulletin*,
889 127(11-12), 1730–1751.
- 890 Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, J., Hillier, M., ...
891 Martin, R. (2014). Next generation three-dimensional geologic modeling and
892 inversion. *Economic Geology*, 18, 261–272.
- 893 Julio, C. (2015). *Conditionnement de la modélisation stochastique 3D des réseaux de*
894 *failles* (Unpublished doctoral dissertation). Université de Lorraine.
- 895 Julio, C., Caumon, G., & Ford, M. (2015a). Impact of the en echelon fault connec-
896 tivity on reservoir flow simulations. *Interpretation*, 3(4), SAC23–SAC34. doi:
897 10.1190/INT-2015-0060.1
- 898 Julio, C., Caumon, G., & Ford, M. (2015b). Sampling the uncertainty associated
899 with segmented normal fault interpretation using a stochastic downscaling
900 method. *Tectonophysics*, 639, 56–67. doi: 10.1016/j.tecto.2014.11.013
- 901 Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- 902 Knuth, D. E. (2005). *The Art of Computer Programming, Volume 4: Generating all*
903 *Combinations and Partitions, Fascicle 3*. Addison-Wesley Professional.
- 904 Lallier, F., Antoine, C., Charreau, J., Caumon, G., & Ruiu, J. (2013). Manage-
905 ment of ambiguities in magnetostratigraphic correlation. *Earth and Planetary*
906 *Science Letters*, 371, 26–36. doi: 10.1016/j.epsl.2013.04.019
- 907 Laurent, G., Caumon, G., Bouziat, A., & Jessell, M. (2013, April). A parametric
908 method to model 3D displacements around faults with volumetric vector fields.
909 *Tectonophysics*, 590, 83–93. doi: 10.1016/j.tecto.2013.01.015
- 910 Lecour, M., Cognot, R., Duvinage, I., Thore, P., & Dulac, J.-C. (2001). Mod-

- 911 elling of stochastic faults and fault networks in a structural uncertainty study.
912 *Petroleum Geoscience*, 7(S), S31–S42. doi: 10.1144/petgeo.7.S.S31
- 913 Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions,
914 and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).
- 915 Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., & Betts, P. G. (2012).
916 Locating and quantifying geological uncertainty in three-dimensional models:
917 Analysis of the Gippsland Basin, southeastern Australia. *Tectonophysics*, 546,
918 10–27. doi: 10.1016/j.tecto.2012.04.007
- 919 Lines, L. R., & Newrick, R. T. (2004). *Fundamentals of Geophysical Interpretation*.
920 Society of Exploration Geophysicists. doi: 10.1190/1.9781560801726
- 921 Mai, P. M., Galis, M., Thingbaijam, K. K. S., Vyas, J. C., & Dunham, E. M. (2017,
922 September). Accounting for Fault Roughness in Pseudo-Dynamic Ground-
923 Motion Simulations. *Pure and Applied Geophysics*, 174(9), 3419–3450. doi:
924 10.1007/s00024-017-1536-8
- 925 Manighetti, I., Caulet, C., Barros, L., Perrin, C., Cappa, F., & Gaudemer, Y.
926 (2015). Generic along-strike segmentation of Afar normal faults, East
927 Africa: Implications on fault growth and stress heterogeneity on seismogenic
928 fault planes. *Geochemistry, Geophysics, Geosystems*, 16(2), 443–467. doi:
929 10.1002/2014GC005691
- 930 Mann, C. J. (1993). Uncertainty in geology. In J. C. Davis & U. C. E. Herzfeld
931 (Eds.), *Computers in geology—25 years of progress* (pp. 241–254). Oxford Uni-
932 versity Press, Inc.
- 933 Manzocchi, T., Heath, A., Childs, C., Telles, I., & Carneiro, M. (2019). Modelling
934 fault zone displacement partitioning for risking across-fault juxtaposition. In
935 *81st eage conference and exhibition 2019*.
- 936 Meisling, K. E., Cobbold, P. R., & Mount, V. S. (2001). Segmentation of an
937 obliquely rifted margin, campos and santos basins, southeastern brazil. *AAPG*
938 *bulletin*, 85(11), 1903–1924.

- 939 Munthe, K., Holden, L., Mostad, P., & Townsend, C. (1994). Modelling sub-seismic
940 fault patterns using a Marked Point Process. In *Ecmor iv-4th european confer-*
941 *ence on the mathematics of oil recovery*. doi: 10.3997/2214-4609.201411151
- 942 Nicol, A., Walsh, J., Childs, C., & Manzocchi, T. (2020). The growth of faults. In
943 *Understanding faults* (pp. 221–255). Elsevier.
- 944 Nixon, C. W., Sanderson, D. J., Dee, S. J., Bull, J. M., Humphreys, R. J., & Swan-
945 son, M. H. (2014). Fault interactions and reactivation within a normal-fault
946 network at Milne Point, Alaska. *AAPG Bulletin*, 98(10), 2081–2107. doi:
947 10.1306/04301413177
- 948 Osypov, K., Yang, Y., Fournier, A., Ivanova, N., Bachrach, R., Yarman, C. E., ...
949 Woodward, M. (2013). Model-uncertainty quantification in seismic tomogra-
950 phy: Method and applications. *Geophysical Prospecting*, 61(6), 1114–1134.
- 951 Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., & Jessell, M. (2018,
952 April). Monte Carlo simulation for uncertainty estimation on structural
953 data in implicit 3-D geological modeling, a guide for disturbance distri-
954 bution selection and parameterization. *Solid Earth*, 9(2), 385–402. doi:
955 10.5194/se-9-385-2018
- 956 Pakyuz-Charrier, E., Jessell, M., Giraud, J., Lindsay, M., & Ogarko, V. (2019).
957 Topological analysis in monte carlo simulation for uncertainty estimation. *Solid*
958 *Earth Discussions*, 2019, 1–37. doi: 10.5194/se-2019-78
- 959 Peacock, D., & Sanderson, D. (1991). Displacements, segment linkage and relay
960 ramps in normal fault zones. *Journal of Structural Geology*, 13(6), 721–733.
961 doi: 10.1016/0191-8141(91)90033-F
- 962 Ragon, T., Sladen, A., & Simons, M. (2018). Accounting for uncertain fault geom-
963 etry in earthquake source inversions-i: theory and simplified application. *Geo-*
964 *physical Journal International*, 214(2), 1174–1190.
- 965 Ramsay, J. G., & Huber, M. I. (1987). *The techniques of modern structural geology:*
966 *Folds and Fractures* (Vol. 2). Academic Press.

- 967 Richards, F. L., Richardson, N. J., Bond, C. E., & Cowgill, M. (2015). Interpretational variability of structural traps: implications for exploration risk and
968 volume uncertainty. *Geological Society, London, Special Publications*, 421(1),
969 7–27. doi: 10.1144/SP421.13
970
- 971 Riesner, M., Durand-Riard, P., Hubbard, J., Plesch, A., & Shaw, J. H. (2017, May). Building Objective 3D Fault Representations in Active Tectonic Settings. *Seismological Research Letters*, 88(3), 831–839. doi: 10.1785/0220160192
972
973
- 974 Rivenæs, J. C., Otterlei, C., Zachariassen, E., Dart, C., & Sjøholm, J. (2005). A 3D stochastic model integrating depth, fault and property uncertainty for planning robust wells, Njord Field, offshore Norway. *Petroleum Geoscience*, 11(1),
975 57–65. doi: 10.1144/1354-079303-612
976
977
- 978 Røe, P., Georgsen, F., & Abrahamsen, P. (2014). An uncertainty model for fault shape and location. *Mathematical Geosciences*, 46(8), 957–969.
979
- 980 Rosenbaum, M. S., & Culshaw, M. G. (2003). Communicating the risks arising from geohazards. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166(2), 261–270. doi: 10.1111/1467-985X.00275
981
982
- 983 Rotevatn, A., Jackson, C. A.-L., Tvedt, A. B., Bell, R. E., & Blækkan, I. (2018, August). How do normal faults grow? *Journal of Structural Geology*. doi: 10.1016/j.jsg.2018.08.005
984
985
- 986 Sanfeliu, A., & Fu, K.-S. (1983). A distance measure between attributed relational graphs for pattern recognition. *IEEE transactions on systems, man, and cybernetics*(3), 353–362.
987
988
- 989 Schaaf, A., & Bond, C. E. (2019). Quantification of uncertainty in 3-d seismic interpretation: implications for deterministic and stochastic geomodelling and machine learning. *Solid earth*.
990
991
- 992 Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, 1(1), 27–64. doi: 10.1016/j.cosrev.2007.05.001
993
- 994 Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface sys-*

- 995 *tems* (Vol. 236). John Wiley & Sons.
- 996 Schneeberger, R., de La Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F.,
997 & Herwegh, M. (2017). Methods and uncertainty estimations of 3-d struc-
998 tural modelling in crystalline rocks: a case study. *Solid Earth*, 8(5), 987. doi:
999 10.5194/se-8-987-2017
- 1000 Seiler, A., Aanonsen, S. I., Evensen, G., & Rivenæs, J. C. (2010). Structural sur-
1001 face uncertainty modeling and updating using the ensemble Kalman filter. *SPE*
1002 *Journal*, 15(04), 1–062. doi: 10.2118/125352-PA
- 1003 Sepúlveda, I., Liu, P. L.-F., Grigoriu, M., & Pritchard, M. (2017, September).
1004 Tsunami hazard assessments with consideration of uncertain earthquake slip
1005 distribution and location: Tsumani Hazard and Uncertain Earthquakes.
1006 *Journal of Geophysical Research: Solid Earth*, 122(9), 7252–7271. doi:
1007 10.1002/2017JB014430
- 1008 Smith, T. F., & Waterman, M. S. (1980). New stratigraphic correlation techniques.
1009 *The Journal of Geology*, 88(4), 451–457.
- 1010 Stoyan, D., & Gloaguen, R. (2011, August). Nucleation and growth of geological
1011 faults. *Nonlinear Processes in Geophysics*, 18(4), 529–536. doi: 10.5194/npg-18
1012 -529-2011
- 1013 Suzuki, S., Caumon, G., & Caers, J. (2008). Dynamic data integration for struc-
1014 tural modeling: model screening approach using a distance-based model
1015 parameterization. *Computational Geosciences*, 12(1), 105–119. doi:
1016 10.1007/s10596-007-9063-9
- 1017 Tal, Y., Hager, B. H., & Ampuero, J. P. (2018, January). The Effects of Fault
1018 Roughness on the Earthquake Nucleation Process. *Journal of Geophysical*
1019 *Research: Solid Earth*, 123(1), 437–456. doi: 10.1002/2017JB014746
- 1020 Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, 2(8),
1021 492–494. doi: 10.1038/nphys375
- 1022 Thiele, S. T., Jessell, M. W., Lindsay, M., Wellmann, J. F., & Pakyuz-Charrier, E.

- 1023 (2016, October). The topology of geology 2: Topological uncertainty. *Journal*
 1024 *of Structural Geology*, *91*, 74–87. doi: 10.1016/j.jsg.2016.08.010
- 1025 Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T., & Cognot, R. (2002). Struc-
 1026 tural uncertainties: Determination, management, and applications. *Geophysics*,
 1027 *67*(3), 840–852. doi: 10.1190/1.1484528
- 1028 Torabi, A., & Berg, S. S. (2011). Scaling of fault attributes: A review. *Marine and*
 1029 *Petroleum Geology*, *28*(8), 1444–1460. doi: 10.1016/j.marpetgeo.2011.04.003
- 1030 Tvedt, A. B., Rotevatn, A., & Jackson, C. A. (2016). Supra-salt normal fault
 1031 growth during the rise and fall of a diapir: Perspectives from 3D seismic reflec-
 1032 tion data, Norwegian North Sea. *Journal of Structural Geology*, *91*, 1–26. doi:
 1033 10.1016/j.jsg.2016.08.001
- 1034 Walsh, J., Nicol, A., & Childs, C. (2002). An alternative model for the growth of
 1035 faults. *Journal of Structural Geology*, *24*(11), 1669–1675. doi: 10.1016/S0191
 1036 -8141(01)00165-1
- 1037 Watterson, J. (1986). Fault dimensions, displacements and growth. *Pure and Applied*
 1038 *Geophysics*, *124*(1-2), 365–373.
- 1039 Wellmann, J. F., & Caumon, G. (2018, Nov). 3-D structural geological models:
 1040 Concepts, methods, and uncertainties. In I. S. C. (Ed) (Ed.), *Advances in geo-*
 1041 *physics* (Vol. 59, p. 1-121). Elsevier. doi: 10.1016/bs.agph.2018.09.001
- 1042 Wellmann, J. F., Horowitz, F. G., Schill, E., & Regenauer-Lieb, K. (2010, July). To-
 1043 wards incorporating uncertainty of structural data in 3D geological inversion.
 1044 *Tectonophysics*, *490*(3-4), 141–151. doi: 10.1016/j.tecto.2010.04.022
- 1045 Wellmann, J. F., Lindsay, M., Poh, J., & Jessell, M. (2014). Validating 3-D struc-
 1046 tural models with geological knowledge for improved uncertainty evaluations.
 1047 *Energy Procedia*, *59*, 374–381. doi: 10.1016/j.egypro.2014.10.391
- 1048 Wilson, C. G., Bond, C. E., & Shipley, T. F. (2019). How can geologic decision mak-
 1049 ing under uncertainty be improved? *Solid earth*, 1–34.
- 1050 Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., & Caumon, G. (2019, Octo-

- 1051 ber). Building realistic structure models to train convolutional neural net-
1052 works for seismic structural interpretation. *GEOPHYSICS*, 1-48. doi:
1053 10.1190/geo2019-0375.1
- 1054 Yielding, G. (2016). The geometry of branch lines. *Geological Society, London, Special*
1055 *Publications*, 439, SP439–1. doi: 10.1144/SP439.1
- 1056 Zakian, P., Khaji, N., & Soltani, M. (2017, October). A Monte Carlo adapted
1057 finite element method for dislocation simulation of faults with uncertain
1058 geometry. *Journal of Earth System Science*, 126(7), 105. doi: 10.1007/
1059 s12040-017-0878-z
- 1060 Zehner, B., Börner, J. H., Görz, I., & Spitzer, K. (2015, June). Workflows
1061 for generating tetrahedral meshes for finite element simulations on com-
1062 plex geological structures. *Computers & Geosciences*, 79, 105-117. doi:
1063 10.1016/j.cageo.2015.02.009
- 1064 Zhu, S., Hack, R., Turner, K., & Hale, M. (2003). How far will uncertainty of the
1065 subsurface limit the sustainability planning of the subsurface? In *Proc. sus-*
1066 *tainable development & management of the subsurface (sdms) conference* (pp.
1067 5–7).