



**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*, 2020, Accepted Manuscript, 54 p. 10.1029/2020JB020022 . hal-02562611v1

**HAL Id: hal-02562611**

**<https://hal.univ-lorraine.fr/hal-02562611v1>**

Submitted on 4 May 2020 (v1), last revised 20 Feb 2021 (v2)

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

3           **Gabriel Godefroy <sup>1</sup>, Guillaume Caumon <sup>1</sup>, Gautier Laurent <sup>1,2</sup>, and François**  
4           **Bonneau <sup>1</sup>**

5                           <sup>1</sup> Université de Lorraine, CNRS, GeoRessources, ENSG

6                           <sup>2</sup> Univ. Orléans, CNRS, BRGM, ISTO, UMR 7327

7                                   <sup>1</sup>F-54000 Nancy, France

8                                   <sup>2</sup>F-45071, Orléans, France

9           **Key Points:**

- 10           • Several plausible scenarios can be made when interpreting faulted structures from  
11           sparse subsurface data.
- 12           • From numerical rules expressing conceptual knowledge, a graph-based sampler gen-  
13           erates several possible fault scenarios honoring spatial data.
- 14           • Numerical experiments suggest that the use of coherent interpretation rules in-  
15           creases the likelihood of generating correct interpretations.

---

Corresponding author: Gabriel Godefroy, [Gabriel.Godefroy@hotmail.fr](mailto: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 proves 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 a possibility graph, whose edges represent the independent association of any two pieces of fault evidence. Each edge carries a likelihood that the endpoints belong to the same fault surface is computed, expressing 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 tedious 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 uncertainty 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.

**1 Introduction****Introduction**

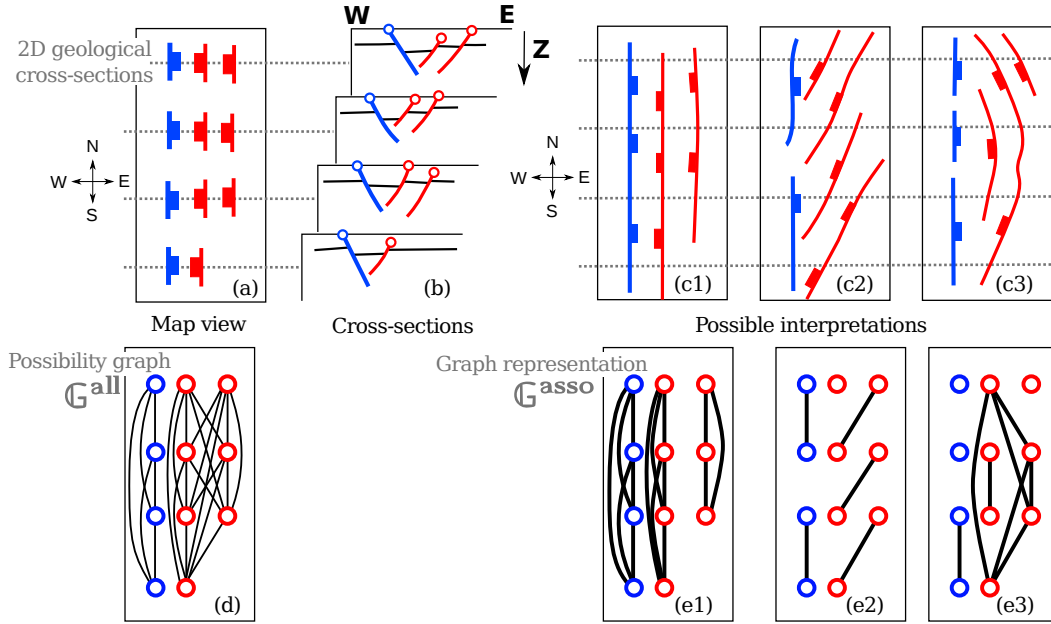
In structural characterization of geophysical data and geological mapping, the lack of conclusive observations generally makes interpretation necessary to obtain a consistent subsurface model. Indeed, geological observations and geophysical signals often have an incomplete spatial coverage and non-unique interpretations due to a lack of resolution or physical ambiguities (Wellmann & Caumon, 2018). As a result, structural uncertainty often remains after interpretation, and it affects fundamental research on earth’s structure such as, for example, the understanding of rift development and earthquake processes (Mai et al., 2017; Riesner et al., 2017; Sepúlveda et al., 2017; Zakian et al., 2017; Gombert et al., 2018; Ragon et al., 2018; Tal et al., 2018). Structural uncertainty also impacts applied studies, for example on natural resource exploration and exploitation (Hollund et al., 2002; Julio et al., 2015a; Richards et al., 2015; Rivenæs et al., 2005; Seiler et al., 2010; Thore et al., 2002), waste disposal (Mann, 1993; Schneeberger et al., 2017),

66 environmental engineering (Rosenbaum & Culshaw, 2003), civil engineering works (Zhu  
67 et al., 2003). In all these applications, faults are often the source of significant geomet-  
68 rical and petrophysical complexity, as fault zones often have distinct hydromechanical  
69 properties, and also because fault displacement controls the geometric layout of rock units  
70 in the subsurface. Understanding and reducing fault uncertainty is, therefore, essential  
71 in many geoscience studies.

72 To characterize structural uncertainty, one may ask a population of geologists to  
73 interpret a particular data set (e.g., Bond et al., 2007; Schaaf & Bond, 2019). However,  
74 interpreting a subsurface data set in three dimensions commonly takes up to several months,  
75 so this strategy is difficult to generalize. Alternatively, one may use computing power  
76 assess uncertainty in structural models. For this, stochastic structural modeling has al-  
77 ready been proposed to generate several scenarios while taking account of seismic im-  
78 age quality and faults below seismic resolution (Aydin & Caers, 2017; Hollund et al., 2002;  
79 Holden et al., 2003; Irving et al., 2010; Julio et al., 2015a, 2015b; Lecour et al., 2001);  
80 uncertainty related to reflection seismic acquisition and processing (Osypov et al., 2013;  
81 Thore et al., 2002); geological field measurement uncertainty (Jessell et al., 2014; Lind-  
82 say et al., 2012; Pakyuz-Charrier et al., 2019; Wellmann et al., 2014); structural param-  
83 eters for folding (Grose et al., 2019, 2018); and observation gaps (Aydin & Caers, 2017;  
84 Cherpeau et al., 2010b; Cherpeau & Caumon, 2015; Holden et al., 2003). Considering  
85 several structural interpretations has also proved useful to propagate uncertainties to flow  
86 simulations (Julio et al., 2015b), to rank structural models against physical data and ul-  
87 timately to falsify some of the interpretations using a Bayesian approach (Cherpeau et  
88 al., 2012; de la Varga & Wellmann, 2016; Irakarama et al., 2019; Seiler et al., 2010; Suzuki  
89 et al., 2008; Tarantola, 2006; Wellmann et al., 2014). In the above approaches, several  
90 conceptual models are used to describe structural uncertainties. A general research ques-  
91 tion is whether the methods produce “geologically realistic” models, what this term ac-  
92 tually means, and how it can be effectively embedded in conceptual models of uncertainty  
93 (Caumon, 2010; de la Varga & Wellmann, 2016; Jessell et al., 2010, 2014; Thibaut et al.,  
94 1996; Wellmann & Caumon, 2018).

95 During the interpretation of faults from sparse observations, the choice of a par-  
96 ticular conceptual model can be consequential to explore and reduce uncertainties. Faults  
97 are typically inferred from observations made on outcrops, wells, geophysical images, or  
98 through the inversion of focal mechanisms. Classically, the interpretation of these data  
99 is translated into points, lines or surfaces indicating the fault position and orientation.  
100 These geometric interpretations (*fault evidence* or *fault data*) may themselves be uncer-  
101 tain (existence, geometry, connectivity). The problem of modeling geologically realistic  
102 fault structures from such incomplete fault data has been initially described by Freeman  
103 et al. (1990) (Figure 1a,b,c), who proposed a methodology based on displacement anal-  
104 ysis to help geologists manually choose between various fault scenarios.

105 Generating these scenarios, however, needs efficient computational techniques to  
106 explore the possibility space. Statistical point processes provide a general mathemati-  
107 cal framework for this (Holden et al., 2003). As tectonic history places specific constraints  
108 on fault networks in terms of orientation and truncation patterns, it is possible to rep-  
109 resent each fault surface as a level set and to sequentially simulate fault sets to repro-  
110 duce specific statistics for each fault set, while enforcing abutting relationships between  
111 the simulated faults (Aydin & Caers, 2017; Cherpeau et al., 2010b, 2012; Cherpeau &  
112 Caumon, 2015). For honoring spatial fault data, Aydin and Caers (2017) use an extended  
113 Metropolis sampler which, at each stage of the simulation, adds, removes, or modifies  
114 a fault object. This sampler has theoretical convergence properties, but simulating fault  
115 networks in the presence of a large number of fault data remains computationally chal-  
116 lenging. Therefore, Cherpeau et al. (2010b, 2012) and Cherpeau and Caumon (2015) pro-  
117 pose a parsimonious method which anchors the first simulated faults to the available ev-  
118 idence, before simulating unseen fault objects. All these iterative stochastic fault mod-



**Figure 1.** Associating labeled pieces of fault evidence (red: east-dipping and blue: west-dipping) interpreted (a) in map view or (b) on two-dimensional seismic lines is an under-constrained problem. (c<sub>1</sub>, c<sub>2</sub> and c<sub>3</sub>) Several structural interpretations are possible (Modified from Freeman et al., 1990; Godefroy et al., 2019). (d) In a possibility graph ( $\mathbb{G}^{\text{all}}$ ), the labeled nodes (i.e., the pieces of fault evidence) are linked by an edge if they may be part of the same fault. (e<sub>1</sub>, e<sub>2</sub> and e<sub>3</sub>) Plausible interpretations are represented by an association graph where the edges link pieces of fault evidence interpreted as belonging to the same fault.

119 els are difficult to use in practice, primarily because of the combinatorial complexity of  
 120 the problem (Godefroy et al., 2019; Julio, 2015), of the difficulty to integrate geological,  
 121 kinematical and mechanical concepts into the stochastic model (Godefroy et al., 2017;  
 122 Laurent et al., 2013; Nicol et al., 2020; Røe et al., 2014; Rotevatn et al., 2018), and of  
 123 the geometric challenges to robustly build such three-dimensional structural models (e.g.,  
 124 due to meshing issues, see Anquez et al., 2019; Zehner et al., 2015).

125 This paper focuses on the combinatorial problem, as it is a prerequisite to address  
 126 the other challenges. For this, we split stochastic structural simulation into three sub-  
 127 problems: (1) The *fault data association problem* (also termed fault correlation by Free-  
 128 man et al., 1990), which aims at determining which of the pieces of evidence may belong  
 129 to the same fault; (2) The interpolation problem, which determines fault geometry and  
 130 displacement from available data, and which has been extensively addressed in deter-  
 131 ministic geological modeling (see Section 3 of Wellmann & Caumon, 2018, and refer-  
 132 ences therein); (3) The simulation of unobserved structures, which can be addressed by  
 133 appropriate statistical point processes (Aydin & Caers, 2017; Cherpeau et al., 2010b; Holden

134 et al., 2003). As items (2) and (3) have already received attention, this contribution fo-  
 135 cuses on finding a suitable computational method for solving the data association prob-  
 136 lem. For this, we build on a recent formalism (Godefroy et al., 2019), where an associ-  
 137 ation scenario is represented by an undirected graph ( $\mathbb{G}^{\text{asso}}$ , Figure 1.e). In this graph,  
 138 each labeled node represents one piece of fault evidence and each edge associates two nodes  
 139 belonging to the same fault. Each connected component represents a fault, and is nec-  
 140 essarily a complete subgraph (or clique) of  $\mathbb{G}^{\text{asso}}$ : all the nodes of a connected compo-  
 141 nent are mutually connected.  $\mathbb{G}^{\text{asso}}$  is a subset of a much larger possibility graph  $\mathbb{G}^{\text{all}}$ ,  
 142 which describes all possible pairwise associations of the available pieces of evidence (Fig-  
 143 ure 1.d). A key idea of Godefroy et al. (2019) is to assign weights to the edges of  $\mathbb{G}^{\text{all}}$ ,  
 144 to represent the likelihood that any pair of fault data belongs to the same fault object  
 145 based on prior geological knowledge. Below a certain likelihood, the edge can be removed,  
 146 which significantly reduces the number of possible scenarios. The analysis of maximal  
 147 cliques in this simplified graph provides fault association scenarios, but the ability of the  
 148 approach to sensibly sample fault uncertainty has not been demonstrated.

149 In the present contribution, we propose a stochastic graph decomposition algorithm  
 150 to automate the generation of several possible fault scenarios (or, equivalently, graphs  
 151  $\mathbb{G}^{\text{asso}}$ ) from an input graph  $\mathbb{G}^{\text{all}}$  representing fault data and structural knowledge (Sec-  
 152 tion 2). A lightweight graph data structure carries the structural interpretation and en-  
 153 ables to generate millions of alternative models. To study the properties of the model  
 154 space sampled by this algorithm, we consider a reference model built from high-resolution  
 155 seismic data, offshore Brazil, and extract from this reference model several sparse data  
 156 sets of variable density. We combine several likelihood criteria translating varying de-  
 157 grees of geological knowledge to check the consistency of the method (Section 3). We then  
 158 discuss how this approach may help address some longstanding challenges for integrat-  
 159 ing data and knowledge and better understand brittle structures in the Earth’s crust (Sec-  
 160 tion 4).

## 161 2 Multi-scenario interpretations using graph decomposition

162 In the graph framework proposed by Godefroy et al. (2019), the simulation of fault  
 163 network scenarios amounts to decomposing the possibility graph  $\mathbb{G}^{\text{all}}$  into a graph  $\mathbb{G}^{\text{asso}}$   
 164 composed solely of disjoint cliques (i.e., fully connected sets of nodes) corresponding to  
 165 fault surfaces. This means that simulating likely fault networks using this formalism is  
 166 equivalent to generating random decompositions of the graph  $\mathbb{G}^{\text{all}}$  into a set of cliques.  
 167 Several graph clustering methods are available in the literature (e.g., Schaeffer, 2007).  
 168 These algorithms generally provide one single decomposition and are thus not directly  
 169 applicable to uncertainty assessment by stochastic simulation. Before explaining the pro-  
 170 posed decomposition algorithm, we first describe how information can be attached to the  
 171 graph nodes and edges to represent geological concepts.

### 172 2.1 Accounting for geological knowledge

173 As the number of fault data for a given area can be very large, we propose to use  
 174 the concept of fault families to reduce the number of possible associations, as done clas-  
 175 sically in fracture and fault analysis (e.g., Cherpeau et al., 2010b; Henza et al., 2011; Nixon  
 176 et al., 2011). Indeed, regional geological knowledge generally includes a description of  
 177 tectonic phases through time, which can be translated in terms of statistical descriptions  
 178 for fault and fracture families. Each piece of fault evidence may be, in this context, la-  
 179 beled by a probability to belong to a particular fault family. Family rules  $R_{\varphi}^{\text{fam}}(\mathbf{v}_i)$  quan-  
 180 tify the likelihood that a fault data  $\mathbf{v}_i$  belongs to the given family denoted by the index  
 181  $\varphi$  (Godefroy et al., 2019). Family rules attach to each piece of evidence a number be-

182 tween 0 (if  $\mathbf{v}_i$  cannot belong to the fault family  $\varphi$ ) and 1 (if it is highly likely that  $\mathbf{v}_i$   
 183 belongs to the fault family  $\varphi$ ). This score is attributed by comparing semantic informa-  
 184 tion about a piece of fault evidence, stored in the form of node labels, and general prior  
 185 information about a given fault family. The node labels may carry kinematic character-  
 186 istics, such as the observed fault type (for example normal or reverse) or the apparent  
 187 orientation of interpreted fault lines.

188 As a simple example, the colors in Figure 1 reflect a deterministic assignment of  
 189 all graph nodes to west-dipping or east-dipping fault families. In the graph formalism,  
 190 this means that the global association graph can be decomposed into several disjoint as-  
 191 sociation subgraphs:

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

192 where  $n$  is the number of fault families ( $n = 2$  in Figure 1.e). Each association scenario  
 193 corresponds in turn to a decomposition of each subgraph  $\mathbf{G}_{\varphi}^{\text{asso}}$  into a set of isolated cliques.  
 194 In the possibility graph  $\mathbf{G}^{\text{all}}$ , the decomposition into family subgraphs  $\mathbf{G}_{\varphi}^{\text{all}}$  may also be  
 195 based on family tags attached to each piece of fault evidence (Figure 1.d). Note that in  
 196 this case, the subgraphs  $\mathbf{G}_{\varphi}^{\text{all}}$  may not always be disjoint. Indeed, a node corresponding  
 197 to a piece of fault evidence may belong to several fault families, hence to several possi-  
 198 bility subgraphs.

199 To further include geological knowledge in the simulator, an association rule  $R_{\varphi}^{\text{assoc}}(\mathbf{v}_i \leftrightarrow$   
 200  $\mathbf{v}_j)$  quantifies the likelihood that two pieces of evidence ( $\mathbf{v}_i$  and  $\mathbf{v}_j$ ) of the same family  
 201 ( $\varphi$ ) belong to the same fault (Godefroy et al., 2019). An association rule  $R_{\varphi}^{\text{assoc}}(\mathbf{v}_i \leftrightarrow$   
 202  $\mathbf{v}_j)$  returns a number between 0 ( $\mathbf{v}_i$  and  $\mathbf{v}_j$  cannot belong to the same fault of the fam-  
 203 ily  $\varphi$ ) and 1 (if both fault data are likely to belong to the same fault). Association rules  
 204 are defined from general structural concepts about faults and from some geometric char-  
 205 acteristics associated with each fault family. For example, it can rely on the distance be-  
 206 tween the evidence, the consistency between the fault family orientation and the orien-  
 207 tation of the edge connecting two pieces of evidence (Cherpeau & Caumon, 2015), the  
 208 throw gradient along the fault (e.g., Barnett et al., 1987; Freeman et al., 1990; Cherpeau  
 209 & Caumon, 2015), or the estimated separation across a faulted area (Freeman et al., 2010).

## 210 2.2 Size of the search space

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

219 In the general case, however, the graph  $\mathbf{G}^{\text{all}}$  is incomplete; it is currently impos-  
 220 sible to perform a more accurate combinatorial analysis just by considering the struc-  
 221 ture of the graph  $\mathbf{G}^{\text{all}}$ . Upper bounds have been proposed for the number of cliques in  
 222 an arbitrary graph (e.g., theorem 3 in Wood, 2007), but not for the number of graph par-  
 223 titions into *disjoint* cliques. This difficulty impacts the evaluation of results produced  
 224 by any uncertainty quantification method, as the size of the search space cannot be eval-  
 225 uated without solving the highly challenging problem of explicitly listing all the possi-  
 226 ble configurations (Knuth, 2005).

227

### 2.3 Sampling by stochastic graph partitioning

228

229

230

231

232

233

234

To generate several scenarios, we propose a hierarchical method relying on sets of nodes that are all connected (cliques) in the possibility graph of at least one family  $\mathbb{G}_\varphi^{\text{all}}$ . In graph theory, cliques that cannot be enlarged without adding new edges are maximal cliques. Maximal cliques can be detected using the Bron-Kerbosch algorithm (Bron & Kerbosch, 1973). Maximal cliques are used at the beginning of the process in order to mimic a manual interpretation where the geologist starts by interpreting the major structures.

235

236

237

238

239

240

241

242

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

243

244

245

246

247

248

**Creation and segmentation of the possibility graph:** the family and association rules are used to compute a graph of all possible associations  $\mathbb{G}_\varphi^{\text{all}}$  for each family  $\varphi$  (Godefroy et al., 2019) (Figure 2, step 2.1). A wide range of formulas can be used to compute the likelihoods according to the defined rules. Each edge  $e^\varphi(\mathbf{v}_i, \mathbf{v}_j)$  of  $\mathbb{G}_\varphi^{\text{all}}$  (linking two pieces of evidence  $\mathbf{v}_i$  and  $\mathbf{v}_j$ ) carries an association likelihood  $L_\varphi^{\text{all}}(\mathbf{v}_i \leftrightarrow \mathbf{v}_j)$  for each family  $\varphi$ :

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

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

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

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

270

271

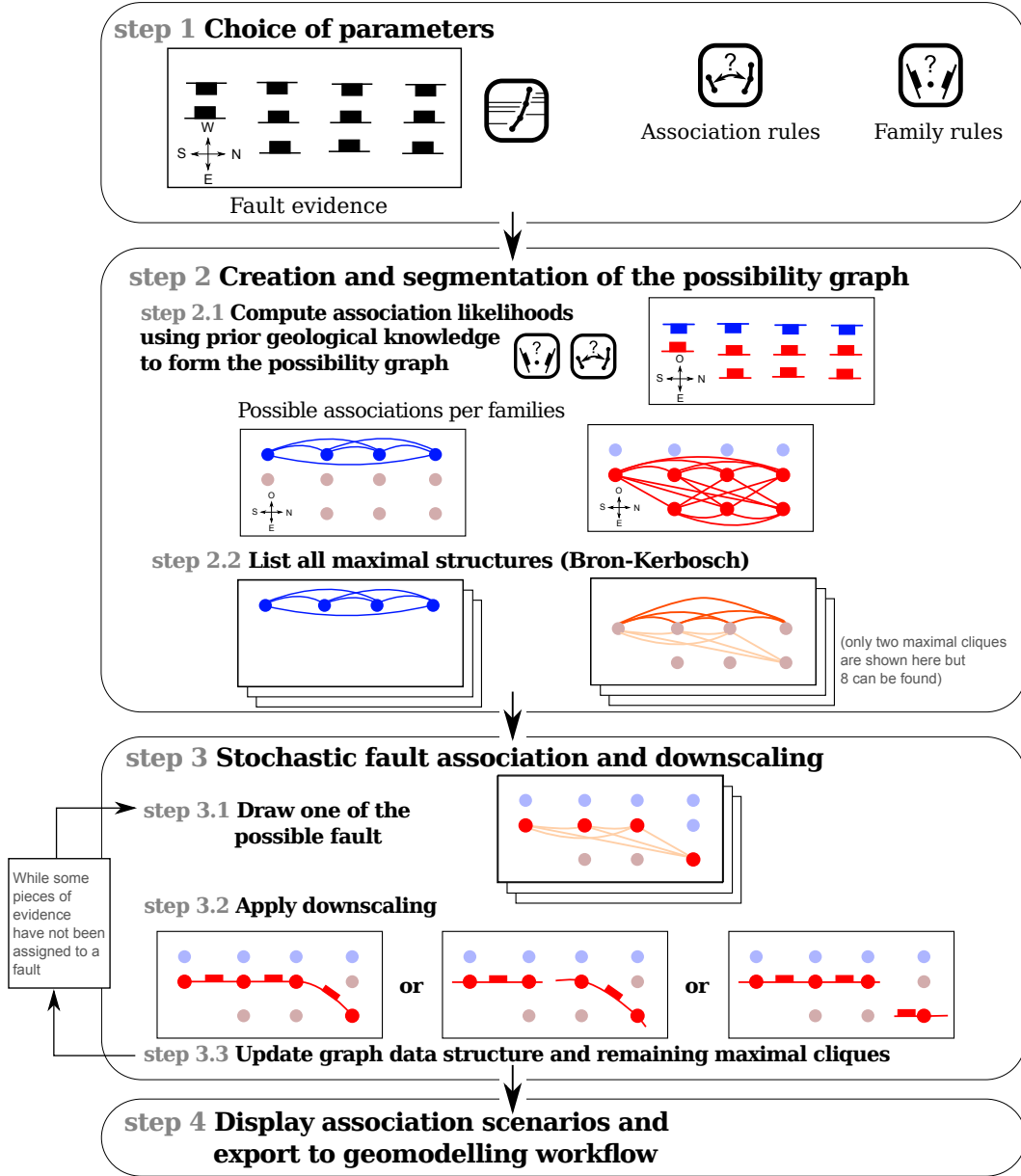
272

273

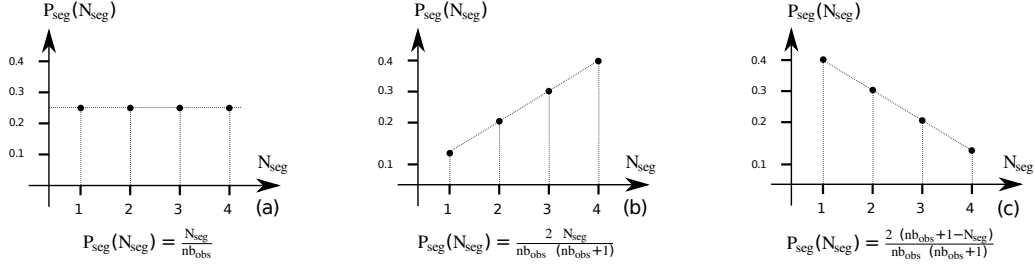
274

where  $\alpha_{\text{draw}}$  is used to weight the number of fault evidences in the clique (the structure containing more nodes are more likely to be drawn when  $\alpha_{\text{draw}}$  increases, see sensitivity study in Section 3.2). Other selection strategies using, for example, the distance separating the pieces of evidence or their sizes could also be used to create large structures.





**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  $\varphi$ . (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.



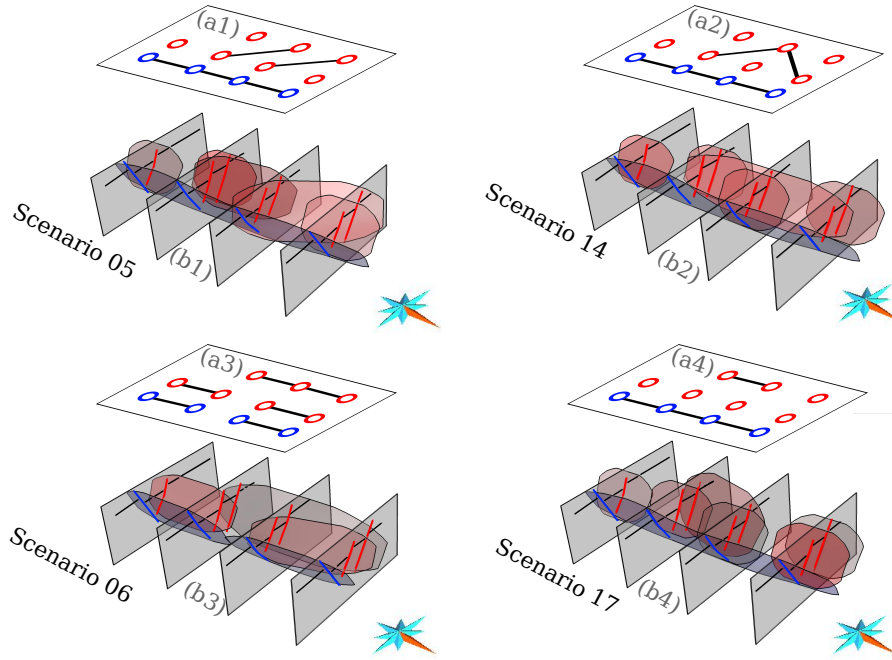
**Figure 3.** Three probability density functions that can be used to draw the number of segments during the segmentation step: (a) uniform, (b) increasing, and (c) decreasing.

275 The maximal clique listing is a way to start the interpretation in a parsimonious  
 276 manner by looking at the major potential structures. However, these potential major  
 277 structures can be made of several faults or several segments possibly linked by  
 278 relay zones (e.g., Ferrill et al., 1999; Julio et al., 2015a; Manighetti et al., 2015;  
 279 Peacock & Sanderson, 1991). These potential fault segments should be considered  
 280 to completely explore the uncertainty space. For this, we propose a simple pro-  
 281 cedure to split a fault  $\mathbb{F}$  made of  $|\mathbb{F}|$  pieces of fault evidence into  $N_{seg}$  fault seg-  
 282 ments; this strategy is called *downscaling* in Julio et al. (2015a, 2015b). For sim-  
 283 plicity, the number  $N_{seg}$  is drawn randomly in this paper between 1 (no segmen-  
 284 tation) and  $|\mathbb{F}|$  (each piece of evidence explains one individual fault segment). This  
 285 random selection relies on a probability law called  $P_{seg}$ , which can be either uni-  
 286 form, linearly decreasing or increasing (Figure 3.a, b, c, respectively). A sensitiv-  
 287 ity analysis is presented in Section 3.2 to show how the choice of  $P_{seg}$  impacts the  
 288 total number of detected structures.

289 In the above simulation method, faults are simulated independently, and the possi-  
 290 bility of interactions is not considered. Indeed, we consider that branch lines are  
 291 not present in the data, as it is very unlikely that branch lines are directly observed  
 292 on typical subsurface data (Yielding, 2016). For this reason, we consider that each  
 293 piece of fault evidence can belong only to one fault surface, so the nodes corre-  
 294 sponding to a selected clique are not considered in further simulation steps (Fig-  
 295 ure 2, step 3.3).

296 As a result of this process, it is possible that crossing faults are generated by the  
 297 simulation; in some geological contexts, it might be reasonable to make the hypo-  
 298 thesis that fault intersections at large scale are unlikely (as in Schneeberger et  
 299 al., 2017). When fault data have been interpreted along parallel two-dimensional  
 300 seismic lines, a labelling is used to detect possible intersections between the sim-  
 301 ulated fault segments and the remaining cliques. After each downscaling step (step 3.2),  
 302 the edges crossing the simulated fault are also removed from the remaining cliques  
 303 while updating the data structure (step 3.3). A similar graph problem is solved  
 304 using the Dynamic Time Warping algorithm (Levenshtein, 1966) in order to cor-  
 305 relate stratigraphic data along wells (Edwards et al., 2018; Lallier et al., 2013; Smith  
 306 & Waterman, 1980). Note that the proposed strategy to avoid intersections is ap-  
 307 plicable only in the presence of fault interpretations made on parallel sections. Ex-  
 308 tending this constraint to irregularly sampled sparse data would require sweep-  
 309 ing the possibility graph to traverse the fault data.

310 **Fault surface modeling:** the presented strategy generates data association scenarios  
 311 that are exported into a visual representation (Figure 2, step 4). Each scenario  
 312 is represented by an association graph where each fault corresponds to a connected



**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 evi-

dence. Data and generated models can be visualized here: [gabrielgodefroy.github.io/](https://gabrielgodefroy.github.io/)

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

313 component of the association graph (Figure 4.a). Fault surfaces can then be in-  
 314 terpolated from the available evidence (e.g., Mallet, 1992). Our implementation  
 315 uses the Structure and Stratigraphy workflow of SKUA-GOCAD (Emerson, 2018).  
 316 This process interpolates each fault surface as the equipotential of a scalar field  
 317 (Frank et al., 2007). The tip line of each fault is classically obtained taking the  
 318 convex hull of the fault data, which may then be manually edited by the interpreter  
 319 (Emerson, 2018). To obtain larger fault extents automatically, we simply compute  
 320 the ellipsoid containing the data (Figures 4.b and supplementary data). As a re-  
 321 finement, fault extents could be sampled from a probability distribution obtained  
 322 by updating the prior distribution for the considered fault family with the avail-  
 323 able information for the current fault data (geometry and observed displacement).  
 324 The geometric uncertainty around the generated association scenario may be fur-  
 325 ther assessed by data perturbation or surface perturbation (see survey in Section  
 326 4 of Wellmann & Caumon, 2018).

### 327 **3 Application to sparse data from Santos basin, offshore Brazil**

328 We applied the proposed stochastic fault network simulation method on a natu-  
 329 ral example of faulted structures imaged by three-dimensional seismic data located in  
 330 the Santos basin, offshore Brasil. The Santos Basin formed during Early Cretaceous when

**Table 1.** Spacing between the 2D lines used to generate the synthetic data.

Number of lines	3	4	5	6
Spacing	2.9 km	1.90 km	1.45 km	1.16 km

331 the South Atlantic began to open. The faults grew from Albian to Miocene and from  
 332 Oligocene to present within Albian Carbonates (Ithahaem Formation) and within Ceno-  
 333 manian to recent fine-grained clastics (Itajai-Acu and Marambia formations). Fault growth  
 334 was activated by the mobilization of an underlying salt-rich unit (Ariri Formation). The  
 335 underlying salt anticlines and salt plateaux can be up to 2.6 km thick (see Tvedt et al.  
 336 (2016) and references therein for more details).

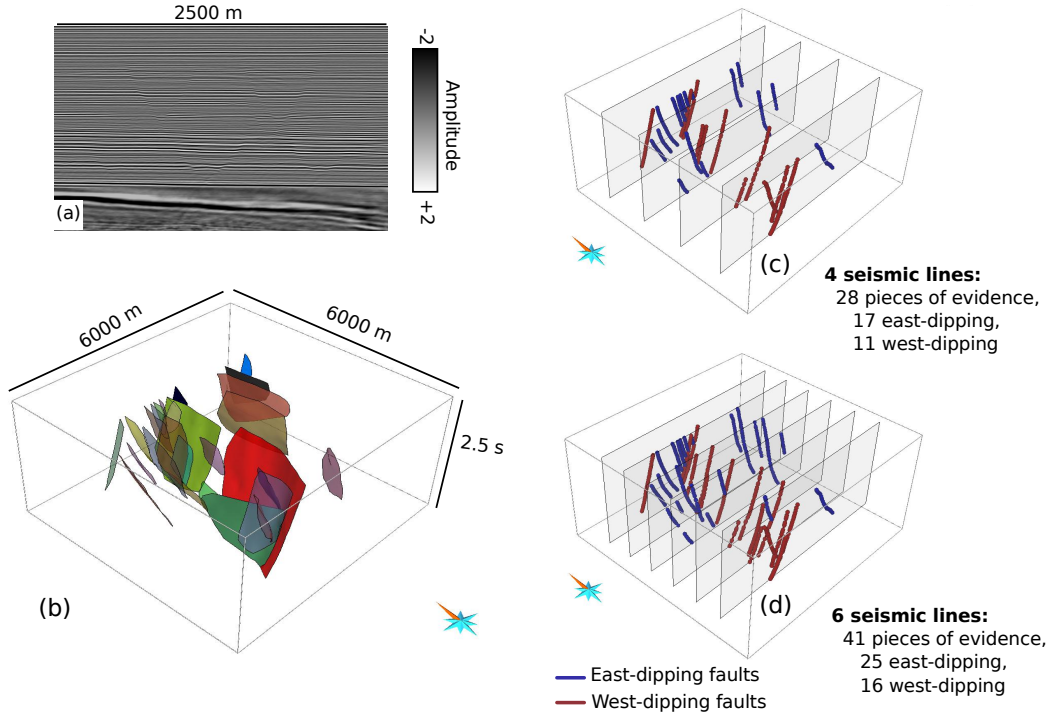
337 From the available time-migrated seismic data (see sample section on Figure 5.a),  
 338 we selected a densely faulted area where we interpreted 27 fault surfaces (see Figure 5.b  
 339 and Godefroy et al. (2017)). Given the excellent quality of the seismic image, there is  
 340 very limited structural uncertainty in this interpretation, which is further used as the  
 341 reference interpretation. From this reference model, we extracted incomplete fault data  
 342 along several parallel two-dimensional sections to emulate the case of the same area be-  
 343 ing imaged by two-dimensional seismic lines (Figure 5.c,d).

344 In the remainder of Section 3, we propose numerical experiments to evaluate the  
 345 consistency of the model space sampled by the proposed simulation method from these  
 346 incomplete data sets. Intuitively, a consistent sampling method should, when appropri-  
 347 ately parameterized, retrieve the reference association with the maximum frequency. Also,  
 348 the likelihood to retrieve the reference association should increase as more data and cor-  
 349 rect informative rules are used. However, in practice, the rules may be biased because  
 350 of preferential sampling or wrong analog knowledge, so we will also check for the impact  
 351 of using biased rules on the ability of the method to find the correct association. Finally,  
 352 in a consistent sampling, the spread of the samples around the reference should also re-  
 353 duce when more information becomes available. However, checking for all these prop-  
 354 erties is difficult in our case, as the dimension of the problem changes when the num-  
 355 ber of observations changes. Therefore, we first study how the structure of the associ-  
 356 ation problem changes with the number of fault data and the degree of information brought  
 357 by geological rules.

### 358 3.1 Synthetic two-dimensional lines and geological rules

359 To quantify the role of a particular geological concept in reducing structural un-  
 360 certainty, we now consider several interpretation rules applied to several data sets of in-  
 361 creasing density extracted from the reference model. The knowledge of the reference model  
 362 enables to determine whether the reference association can be retrieved, and to study  
 363 the influence of chosen geological rules and algorithm parameters on the quality of the  
 364 generated interpretations.

365 The multi-scenario association strategy was applied on fault evidence extracted along  
 366 3, 4, 5, and 6 cross-sections, using a set of geological rules consistent with the reference  
 367 model. The distances between two cross-sections are given Table 1 Because such an ideal  
 368 case is unrealistic in actual sparse data settings, we delete and modify some of the rules  
 369 to test how the rule choices impact structural interpretation.



**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 viewer is available here: [gabrielgodefroy.github.io/StochasticInterpData/Fig5/Fig5.html](https://gabrielgodefroy.github.io/StochasticInterpData/Fig5/Fig5.html)

**Table 2.** Numerical values used in the association rules.

Fault families	Distance		Strike			
	Minimum	Maximum	Minimum	Maximum	Average	Tolerance
Family 1	0 km	3.6 km	N330	N015	N352.5	$\pm 22.5$
Family 2	0 km	3.6 km	N145	N195	N170	$\pm 25$

**Table 3.** 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 3.1. 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.

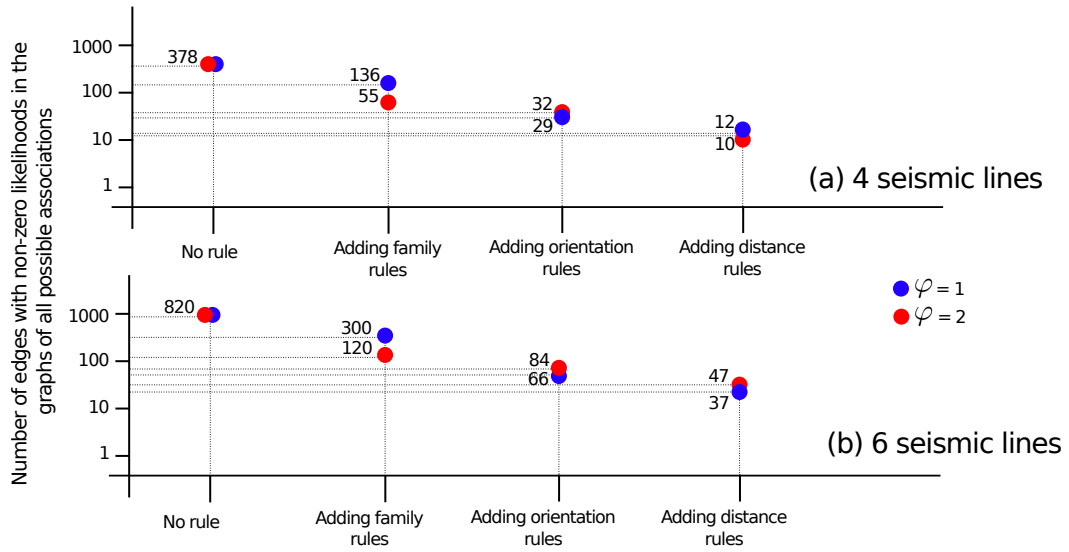
Number of synthetic seismic lines	Number of pieces of evidences...	... in fault family 1	... in fault family 2	Bell numbers	Time to simulate $5.10^7$ realizations
3	25	16	9	$B_{25} = 4, 6.10^{18}$	03h49m
4	28	17	11	$B_{28} = 6, 1.10^{21}$	03h51m
5	35	22	13	$B_{35} = 2, 8.10^{29}$	05h30m
6	41	25	16	$B_{41} = 2, 3.10^{36}$	06h54m

370 In the considered data set, faults have approximately a north/south strike and can  
 371 be grouped into two fault families: east- and west-dipping faults. Two fault family rules  
 372 are defined based on the dip direction. As seismic lines are oriented east/west, the slope  
 373 of fault interpretations completely determines the family, so there is no uncertainty about  
 374 which family each piece of evidence belongs to. To evaluate the likelihood of associat-  
 375 ing two fault traces, we first use an association rule that restricts the strike of the gen-  
 376 erated faults to be between  $N330$  and  $N015$  for the first family, and between  $N145$  and  
 377  $N195$  for the second. Additionally, a uniform association distance rule is created using  
 378 the largest fault extension observed in the reference model (3.6 km). The numerical val-  
 379 ues and the mathematical expressions defining these rules are given respectively in Ta-  
 380 ble 2 and Appendix 6.1.

381 Before applying the simulation algorithm described in Section 2.3, we consider each  
 382 association rule separately to assess the impact of conceptual information on the struc-  
 383 ture of the possibility graph (Figure 6). As expected, these restrictive rules decrease the  
 384 number of edges in  $\mathbb{G}_\varphi^{\text{all}}$ , hence the number of possible association scenarios for each fam-  
 385 ily (see Section 2.2). In the interpreted area of interest, faults do not intersect each other,  
 386 so intersections are also forbidden during the simulations. In spite of these rules, going  
 387 from 4 to 6 seismic sections significantly increases both the number of graph nodes (from  
 388 28 to 41, see Table 3 for details) and graph edges (from 378 to 820), making it more dif-  
 389 ficult to explore the search space or to find the optimal fault configuration.

### 390 3.2 Sensitivity to scale parameters

391 The simulation process is parameterized by a scalar value  $\alpha_{draw}$  and by a proba-  
 392 bility density function  $P_{seg}$ , which both relate to the fault size and impact the number  
 393 of simulated fault surfaces. Figure 7 exhibits statistics on the number of simulated faults,  
 394 while modifying these two parameters. Statistics are computed over  $5.10^5$  realizations,



**Figure 6.** Number of edges per fault family ( $\varphi$ ) 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 possibility graphs can be interactively visualized here: [gabrielgodefroy.github.io/StochasticInterpData/Fig6/html/Fig6.html](https://gabrielgodefroy.github.io/StochasticInterpData/Fig6/html/Fig6.html)

395 with pieces of fault evidence extracted from the Santos Basin model, along 4 and 6 vir-  
 396 tual cross-sections. All of the rules described in Section 3.1 are used for these experiments.

397 The probability density function  $P_{seg}$  used to downscale a fault into several seg-  
 398 ments (Figure 3) clearly influences the mean number of simulated faults (Figure 7a<sub>2</sub>, b<sub>2</sub>).  
 399 As expected, the use of a linearly increasing density function leads to the simulation of  
 400 more faults as compared to the decreasing function. When the number of fault obser-  
 401 vations increases, this trend becomes significant whereas the variability over the total  
 402 set of realizations increases. For applications which seek to preferentially generate par-  
 403 simonious scenarios in terms of number of faults, the choice of a decreasing law seems  
 404 appropriate. In general, however, it seems more relevant to explore the search space, so  
 405 we will choose the uniform law for  $P_{seg}$  from now on, as it almost spans the same ex-  
 406 treme number of faults as with the increasing or decreasing laws.

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

### 417 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 generat-  
 ing 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  $\mathbb{G}^{asso}_l$  and  $\mathbb{G}^{asso}_m$ .  $N_{diff}$  is defined as a spe-  
 cial case of graph edit distance (Sanfeliu & Fu, 1983), in which the only edit operations  
 are edge insertion and deletion:

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

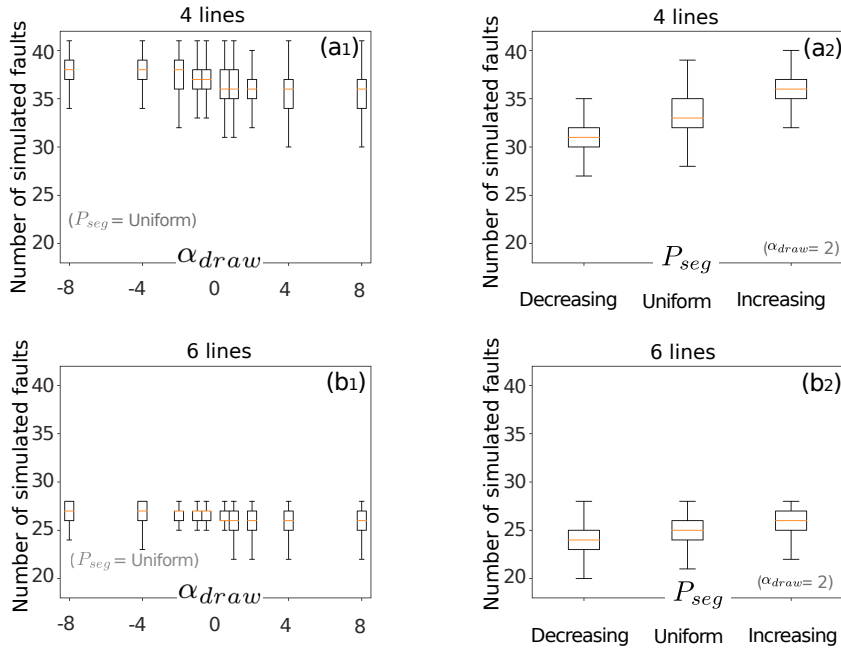
where

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

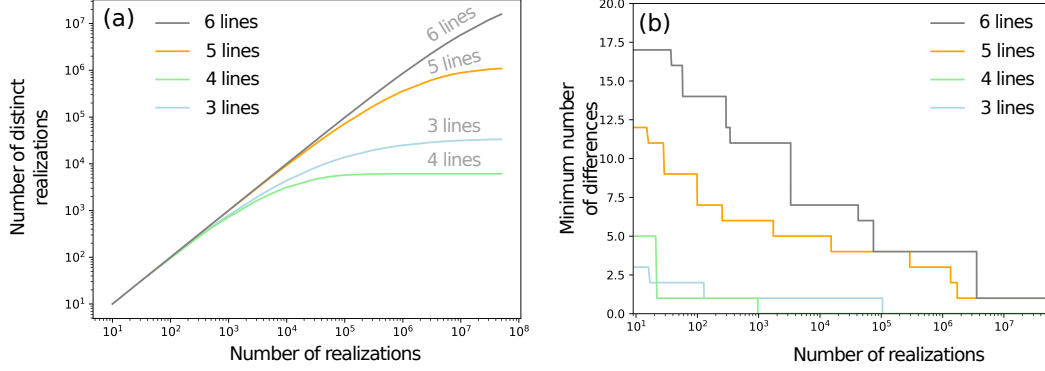
418 For computational reasons,  $5 \cdot 10^7$  realizations were simulated from the fault inter-  
 419 pretations extracted from 3, 4, 5, and 6 virtual seismic lines, using all the previously de-  
 420 scribed rules (Figure 8.a). Simulations run in 3 to 7 hours (see Table 3 for details). At  
 421 the beginning of the simulation, the sampling algorithm shows a near-optimal exploration  
 422 efficiency as it generates only different realizations, whatever the data density.

423 For the case with 3 seismic lines, the plateau is not yet fully reached but for the  
 424 case with 4 seismic lines, a plateau of 6072 distinct realizations is reached after  $3 \cdot 10^6$  re-  
 425 alizations. In both cases, the reference association is found 2026 times and 3164 times  
 426 for 3 and 4 seismic lines, respectively. This can be explained by the different informa-  
 427 tion content carried by these data sets (see Figure 6 of the supplementary material). When  
 428 simulating from 5 and 6 seismic lines (35 and 41 fault data, respectively), the numbers





**Figure 7.** Sensitivity of the number of simulated faults to the parameter  $\alpha_{draw}$  ( $a_1, b_1$ ) and to the probability density function ( $a_2, b_2$ ) used while drawing the number of fault segments  $P_{seg}$ . Statistics computed over  $5 \cdot 10^5$  realizations for the cases with 4 ( $a_1, a_2$ ) and 6 ( $b_1, b_2$ ) seismic lines.



**Figure 8.** (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.

of distinct scenarios is still significantly increasing after  $5.10^7$  realizations and the reference association is not found. In both cases, the best scenario produced has a number of differences  $N_{diff}$  equal to 1 (Figure 8.b).

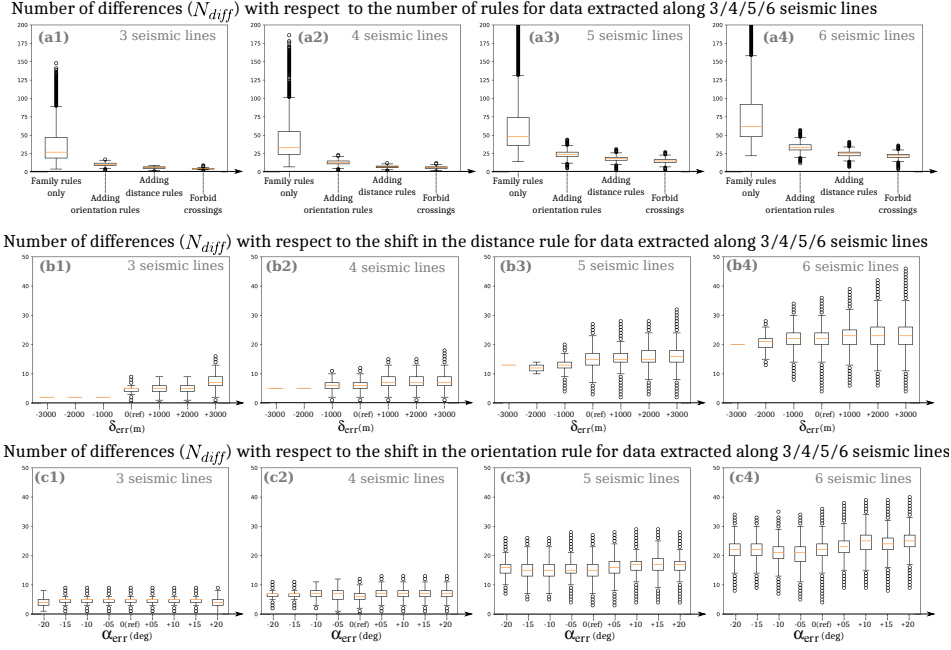
This numerical experiment shows the difficulty of retrieving the reference association when the number of pieces of fault evidence is high, even if the chosen rules are restrictive and consistent. Indeed, as discussed in Section 2.2, the combinatorial complexity increases in a non-polynomial way with the number of nodes. The proposed formalism is thus favorable when relatively few pieces of fault evidence are available, and when the association rules strongly constrain the solution space. This is consistent with expectation that more data brings new knowledge and thus reduces the uncertainties. However, the experiment shows that this information effect can be significantly counterbalanced by the difficulty to explore a larger search space, as also observed by Edwards et al. (2018) for well correlation.

### 3.4 Influence of the chosen geological rules on simulation results

#### 3.4.1 Impact of the number of rules

We tested the impact of the chosen numerical rules by successively running simulations with an increasing number of rules (orientation rule, distance between pieces of fault evidence, and forbidding fault intersections).  $5.10^5$  realizations were generated from digitized fault evidence extracted along the 3, 4, 5, and 6 seismic lines.

To analyze these results, we now consider the number of differences  $N_{diff}(l, ref)$  between each simulated association  $\mathbf{G}_{asso}^l$  and the reference association  $\mathbf{G}_{ref}^{asso}$  interpreted from the full 3D seismic data set. When the simulations are run with more rules (Figure 9.a<sub>1</sub>–a<sub>4</sub>), the minimum, mean, and maximum number of differences  $N_{diff}(l, ref)$  consistently decrease. Moreover, we observe that rules interact with the data density on



**Figure 9.** 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  $5 \cdot 10^5$  realizations.

453 two respects. First, considering a single rule in sparse data settings (cases with 3 or 4  
 454 sections) yields realizations closer to the reference than considering more informative rules  
 455 in denser data settings (cases with 5 or more sections). This can be explained by the non-  
 456 convergence of the sampler for 5 and 6 sections, as shown in Figure 8.a. Second, for a  
 457 given set of rules, the absolute number of differences from the reference increases with  
 458 the number of data. This is a direct effect of the increasing complexity of the search space,  
 459 and may also be explained by the non-convergence of the sampler with  $5 \cdot 10^7$  realizations  
 460 for more than 5 seismic lines (Figure 9.a<sub>4</sub>).

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

### 466 3.4.2 *Deliberately selecting biased rules*

467 In practical geological studies, it would be difficult to come up with appropriate  
 468 parameters for the association rules. Therefore, we tested the impact of choosing erro-  
 469 neous geological concepts during structural interpretation: the orientations defining the  
 470 associations rules were shifted by an angle  $\alpha_{err}$  and the distance  $max_{dist}$  defining the  
 471 distance association rule was offset by  $\delta_{err}$ . The mathematical expressions for the fal-  
 472 sified rules are given in Appendix 6.1.

473 We study the mean number of differences to the reference association (over  $5 \cdot 10^5$   
 474 realizations) according to these inappropriate choices (Figure 9.b,c). Strongly biased rules  
 475 make it impossible to retrieve the reference association from fault evidence sampled along  
 476 3, 4, 5 or 6 seismic lines (Figure 9.b<sub>1</sub>, b<sub>2</sub>, c<sub>1</sub>, c<sub>2</sub>, b<sub>3</sub>, b<sub>4</sub> and c<sub>3</sub>, c<sub>4</sub>, respectively). When sim-  
 477 ulating interpretation scenarios from data extracted along 5 or 6 seismic sections, the  
 478 reference scenario is never retrieved.

479 In the case of the distance rule, if  $\delta_{err}$  is negative, no association is allowed, yield-  
 480 ing a collapse of the ensemble of associations. Such a collapse clearly highlights an in-  
 481 consistency between the rules and the spacing of the synthetic data. On the other hand,  
 482 if the distance rule is more permissive than the reference one (i.e.,  $\delta_{err}$  is positive), then  
 483 the minimum, mean, and maximum numbers of error increase for all number of seismic  
 484 lines.

485 The effect of changing the orientation rule is not as dramatic, but for the cases with  
 486 5 and 6 seismic lines, the deviation from the reference rules leads to a slight increase of  
 487 the mean numbers of error.

## 488 4 Discussion and ways forward

### 489 4.1 On objectivity and uncertainty

490 The proposed graph-based sampling method is a new framework to rapidly and au-  
 491 tomatically explore fault network uncertainties by generating stochastic fault scenarios  
 492 from sparse observations. As compared to classical fault interpretation approaches where  
 493 experts associate the pieces of evidence based on their prior knowledge, our approach  
 494 forces the interpreter to explicitly formulate elementary association rules, which are ag-  
 495 gregated by a stochastic algorithm. The promise of this work is to make interpretation  
 496 an objective, unbiased and reproducible process. However, the geological interpretation  
 497 of subsurface data is precisely about adding some conceptual knowledge to the data, which

498 cannot be fully objective and will always depend on the current state of knowledge and  
 499 experience of an interpreter or on some model assumptions (Bond et al., 2007; Cham-  
 500 berlin, 1890; Frodeman, 1995; Wellmann & Caumon, 2018). In this paper, we used rel-  
 501 atively general rules to define fault data association likelihoods, but the order by which  
 502 we processed and combined them is certainly subjective and driven in part by mathe-  
 503 matical and algorithmic convenience, so we do not claim it to be fully objective. Nonethe-  
 504 less, as compared to classical expert-based interpretation, we see the general approach  
 505 proposed in this paper as a step towards making the interpretation process more trans-  
 506 parent and reproducible using a probabilistic way of expressing and combining geolog-  
 507 ical concepts.

508 Another major difference between this approach and the expert-based method is  
 509 the intrinsic ability of the former to generate several scenarios, whereas most of the lat-  
 510 ter tend to end up with one deterministic solution. A reason is that interpretation ex-  
 511 ercises are taught as a deterministic activity in the vast majority of university courses:  
 512 the general expectation, in surface or subsurface mapping, is to produce only the most  
 513 likely scenario or model. This is shown even in experiments assessing the interpretation  
 514 uncertainty, which ask a set of geoscientists to produce one interpretation each (Bond  
 515 et al., 2007; Bond, 2015; Schaaf & Bond, 2019). Cognitive biases also explain the dif-  
 516 ficulty of one to work with multiple hypotheses (Chamberlin, 1890; Wilson et al., 2019),  
 517 which is a possible explanation for interpretation bias (Bond et al., 2007). The advent  
 518 of computer-based methods makes it easier to explore aleatory uncertainties by perturb-  
 519 ing a reference model (see Wellmann & Caumon, 2018, and references therein), but ad-  
 520 dressing epistemic uncertainties is more challenging as it requires to formalise the geo-  
 521 logical concepts. The method proposed in this paper clearly belongs to this latter class  
 522 of methods.

## 523 4.2 On graphs for structural uncertainty assessment

524 We see the graph-based method proposed in this paper as a possible way to com-  
 525 plement or generalize previous structural uncertainty evaluation approaches:

- 526 • It is a parsimonious approach which starts from existing observations. As such,  
 527 it follows the same philosophy as data perturbation strategies, which consider that  
 528 spatial observations have uncertain location and/or orientation (Lindsay et al., 2012;  
 529 Pakyuz-Charrier et al., 2018; Wellmann et al., 2010). However, a major concep-  
 530 tual difference exists. Indeed, even though the topology of the geological model  
 531 may change by sampling data orientation only, fault data must be associated a  
 532 priori with a particular fault surface. In contrast, our approach starts without any  
 533 particular assumption about how to associate incomplete fault observations to-  
 534 gether. It can be seen as a way to randomly change the data labels, and can be  
 535 used before sampling from orientation or location distributions associated with in-  
 536 complete observations. Data uncertainty can also be integrated in the graph-based  
 537 framework.
- 538 • Another class of methods to address structural uncertainty is to proceed by ge-  
 539 ometric perturbation techniques of an existing structural model (Holden et al., 2003;  
 540 Lecour et al., 2001; Røe et al., 2014). These methods keep the fault connectivity  
 541 of the initial structural model constant, but may change the connectivity of rock  
 542 units on either sides of faults by changing the fault throw. In this paper, we tried  
 543 to reduce the risk of bias by focusing on large-scale fault topological changes, start-  
 544 ing from the observations rather than from an initial fault network interpretation.  
 545 Note, however, that all data known to belong to the same fault surface can be as-  
 546 sociated in a deterministic way in the proposed graph-based method.
- 547 • Approaches addressing topological fault network uncertainty have been proposed  
 548 before, using mainly data-driven iterative simulation methods (Cherpeau et al.,

2010a; Cherpeau & Caumon, 2015; Julio et al., 2015b) or object simulation based on stochastic point processes (Aydin & Caers, 2017; Cherpeau et al., 2010b; Holund et al., 2002; Munthe et al., 1994). All data-driven approaches, including the method presented herein, can be seen as an efficient way to honor observations, which is a notably difficult and time-consuming process when the spacing between observations is smaller than the size of the simulated objects. Data-driven fault simulations are also parsimonious as they only focus on explaining observations, but they can significantly under-estimate the number of faults in a given domain. For example, the Santos case study clearly shows that the number of simulated faults decreases when less data is used for the same area of interest (Figure 7). Therefore, we firmly believe that a stochastic point process should ultimately complement the proposed approach to simulate faults that are not directly supported by observations (Aydin & Caers, 2017; Bonneau et al., 2016; Cherpeau et al., 2010b; Davy et al., 2013; Holden et al., 2003; Munthe et al., 1994; Stoyan & Gloaguen, 2011). Such a process would generate new nodes in the graph  $\mathbb{G}^{\text{asso}}$ . A particular challenge would be to choose an appropriate point process and to come up with a sound parameter inference strategy for the stochastic process. This includes avoiding collisions between simulated faults and existing cross-sections or boreholes where no fault has been observed. Another line of progress in the graph-based method concerns the management fault branch lines. Indeed, even though the chosen reference data set (Figure 5) is free of fault branchings, most faults branch or interact together during their growth (Nicol et al., 2020). A second oriented graph could be considered and updated during the simulation algorithm to represent how faults branch in the fault network. The simulation process should account for the chronology of the development of the successive fault families to preserve the spatial dependency of fault geometry, hierarchy and the fault abutting relationship (as in Aydin & Caers, 2017; Cherpeau et al., 2010b). This would imply updating the association likelihoods (calculated at step 2.1 in Fig. 2) after each stochastic step of the simulation algorithm (step 3.3 of Fig. 2).

Unlike previous iterative methods (Aydin & Caers, 2017; Cherpeau et al., 2010a; Cherpeau & Caumon, 2015), the potential major fault structures are processed in the early steps of our algorithm thanks to the maximal clique detection. Assuming that largest faults are most likely to correspond to many graph nodes, this reproduces a classical interpretation process whereby geologists focus on largest structures before focusing on smaller objects (e.g., Lines & Newrick, 2004). Therefore, we believe that the graph-based approach more effectively explores the search space than previous iterative methods, which tend to proceed by local propagation of information. The proposed clique subdivision is another way to formalize fault downscaling approaches (Julio et al., 2015b; Manighetti et al., 2015). It would be interesting, nonetheless, to quantitatively study how much the proposed strategy affects the sampled model space as compared to previous iterative methods.

### 4.3 Are the produced interpretations “geologically realistic”?

The proposed graph-based method enables the interactive definition of new rules according to a specific geological context. Overall, we only tested relatively simple geological rules in this paper in order to assess the consistency of the sampling algorithm. The method allows for formulating additional rules to improve the consistency of the obtained results and reduce the search space, for example: to choose fault surface orientations from analog data sets (Aydin & Caers, 2017), to account for fault curvature and lateral extension, and an estimation of the fault slip (Cherpeau & Caumon, 2015; Røe et al., 2014), to evaluate fault segmentation (Julio et al., 2015a; Manighetti et al., 2015; Manzocchi et al., 2019).

Such formalizations of geological knowledge as additional numerical rules clearly call for more studies to come up with appropriate choice of rules and parameters. In this paper, we started by choosing an ideal case where the parameters of simple rules were calibrated directly on a reference model. We acknowledge that this is never the case in practice where no reference model exists. Tests made in Section 3.4.2 suggest that, when rule parameters are chosen inappropriately, the simulated models can show significant bias, and we can expect that this observation would also hold for more complex rules. Therefore, further studies are needed to help geoscientists to define the numerical interpretation rules to be used in a particular geological context, how to choose their parameters and how to combine them. This could rely on outcrop or subsurface analog data bases, analog laboratory models, or process-based numerical models, or on an inference process applied directly to the data at hand. Machine learning could also come into play in this process, either by inferring rule parameters or a posteriori assessing the likelihood of the various realizations produced by the sampling method. Training for these approaches could be achieved on multiple manual interpretations (e.g., Schaaf & Bond, 2019) or on processed synthetic models (Wu et al., 2019).

The graph-based sampler, when appropriately parameterized, could, in principle, (1) lead to a more objective characterization of structural uncertainty as compared to manual interpretation and (2) help reducing cognitive biases (Wilson et al., 2019). However, an automatic method such as the graph-based sampling proposed in this paper has no guarantee to produce the same results as several interpretations made by several experts. One reason is that experts can be biased, but this could be addressed in principle by selecting a sufficiently large number of experts. Another, more serious reason, is that the sampling method is likely to miss some important aspects of geological interpretation. On this regard, we see two main avenues for improvement:

- First, the methodology does not completely automate the three-dimensional structural modeling, which makes it difficult to assess the likelihood of the generated fault networks using advanced structural analyzes such as global displacement analysis (Freeman et al., 2010) or structural restoration (Gratier & Guillier, 1993). For this, faults geometries should be modeled using explicit surfaces (Lecour et al., 2001; Røe et al., 2014) or implicit surfaces (Aydin & Caers, 2017; Cherpeau et al., 2010b). Then, the geological formations affected by the fault network should be modeled as done for example by Cherpeau and Caumon (2015); Godefroy et al. (2017); Laurent et al. (2013). The fault geometries need to be simulated while accounting for theoretical tip-line geometries (e.g., Barnett et al., 1987) and theoretical scaling laws (e.g., Torabi & Berg, 2011). Generating such geometries would be useful to assess the impact of fault network uncertainty on resource assessment (Richards et al., 2015), to incorporate this source of uncertainty in geophysical inverse problems (Giraud et al., 2019; Ragon et al., 2018), or to formulate the “geological realism” problem as an Bayesian inference problem (de la Varga & Wellmann, 2016).
- Second, the graph formalism at this stage only considers pairwise associations but does not use the likelihood of associating several nodes at once. For example, it could be likely to associate the pieces of evidence A and B, B and C, and A and C independently, but very unlikely to associate A, B and C altogether. This calls for the definition of multi-point, higher-order statistical rules which apply to graph cliques. For example, one could consider the throw distribution along fault strike (as in Cherpeau & Caumon, 2015; Freeman et al., 1990) or statistical relationships between observed separations and fault size (e.g., Gillespie et al., 1992; Torabi & Berg, 2011). The association likelihoods could also be updated during the graph-based sampling algorithm to account for the fact that fault-fault intersections at large scale are unlikely but are not necessarily impossible (Schneeberger et al., 2017).

To bridge the gap between automatic interpretation methods such as proposed in this paper and classical expert-based techniques, a possible avenue could be to share and interactively render three-dimensional objects on a web server to quantitatively study how a large number of geologists interpret available data. This was recently proposed by Schaaf and Bond (2019) using three-dimensional subsurface models created by a class of students. Using numerical interpretation tools could further enable to record the interpreter strategies, and for example, to detect which geological structures geologists interpret first.

#### 4.4 Inverse problem and clustering of structural interpretations

Stochastic structural modeling enables to generate large numbers of alternative scenarios (several millions in this work) which can be used as prior information in subsurface inverse problems (see Wellmann and Caumon (2018) and references therein). However, the automatic generation of a three-dimensional meshed model for each interpretation is currently impossible. The computational times of focal mechanism inversion, flow simulation, or seismic forward modeling are often incompatible with more than hundreds or maybe thousands of models. Furthermore, for a human being, it seems difficult to work with more than a few alternative scenarios deemed representative of the uncertainties. An effective way to address this problem is to use model clustering in model space (e.g., Suzuki et al., 2008) or in data space (i.e., based on the similarity between their forward response, see for example Scheidt et al., 2018; Irakarama et al., 2019). A challenge, in both cases, comes from the redundancy of models sampled by a particular stochastic methods: indeed, simulation methods tend to generate many similar models in a priori likely regions of the search space. This redundancy is needed if the models also have a large posterior probability, but it can raise efficiency problems when the Bayesian updating is strong. Also, the above methods tend to search for model clusters *after* the sampling stage, which is not optimal.

The graph-based sampler described in this paper opens some avenues to make progress in this area. Indeed, maximal cliques are detected and processed sequentially within the sampling algorithm. Therefore, a hierarchical clustering of structural scenarios could be generated by applying the method in a recursive manner. A possible and simplified outline of such a hierarchical sampling reads:

1. Define the starting possibility graph  $\mathbb{G}_0^{\text{all}} \leftarrow \mathbb{G}^{\text{all}}$  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 possibility graph:  $\mathbb{G}_h^{\text{all}} \leftarrow \mathbb{G}_{h-1}^{\text{all}} \setminus \mathbb{V}_{n_h}$
  - (c) If  $\mathbb{G}_h^{\text{all}}$  still has vertices, go to Step 2, otherwise terminate.

## 5 Conclusions

The proposed graph-based framework helps interpreting alternative fault scenarios to account for the uncertainty arising while considering sparse fault sample. Prior geological knowledge is formalized using numerical geological rules, enforcing the geological consistency of the interpretation choices. The mathematical format of the rules eases the communication of the geological concepts used during the interpretation as compared to manual interpretation approaches, and makes the structural interpretation process reproducible.



Each scenario is represented by a graph. The automatic interpretation framework relies on maximal cliques, i.e., major possible structures in the graphs of all possible associations. This strategy mimics the behavior of an interpreter who would start by the structures explained by a large amount of evidence. The use of a light graph data structure, as compared with a full three-dimensional model, leads to a fast simulation process. This rapidity allows to perform sensitivity studies on the numerical rules and simulation parameters using sparse data extracted from a reference model.

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

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

## Acknowledgments

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

## 6 Appendices

### 6.1 Rules expression

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

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

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

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

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

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

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

736

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

737

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

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

738

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

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

739

with  $\min_{\varphi_1}^{\text{strike}} = 330$ ,  $\max_{\varphi_1}^{\text{strike}} = 15$ ,  $\min_{\varphi_2}^{\text{strike}} = 145$ , and  $\max_{\varphi_2}^{\text{strike}} = 195$ .

740

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

741

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

with  $\max_{\text{dist}} = 3600$  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:

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

742

## References

743

Anquez, P., Pellerin, J., Irakarama, M., Cupillard, P., Lévy, B., & Caumon, G.

744

(2019, January). Automatic correction and simplification of geological maps

745

and cross-sections for numerical simulations. *Comptes Rendus Geoscience*,

746

351, 48-58. doi: 10.1016/j.crte.2018.12.001

747

Aydin, O., & Caers, J. K. (2017). Quantifying structural uncertainty on fault

748

networks using a marked point process within a Bayesian framework. *Tectono-*

749

*physics*, 712, 101–124. doi: 10.1016/j.tecto.2017.04.027

750

Barnett, J. A., Mortimer, J., Rippon, J. H., Walsh, J. J., & Watterson, J. (1987).

751

Displacement geometry in the volume containing a single normal fault. *AAPG*

752

*Bulletin*, 71(8), 925–937.

- 753 Bond, C. (2015). Uncertainty in structural interpretation: Lessons to be learnt.  
754 *Journal of Structural Geology*, *74*, 185–200. doi: 10.1016/j.jsg.2015.03.003
- 755 Bond, C., Gibbs, A., Shipton, Z., & Jones, S. (2007). What do you think this is?  
756 “Conceptual uncertainty” in geoscience interpretation. *GSA Today*, *17*(11), 4.  
757 doi: 10.1130/GSAT01711A.1
- 758 Bonneau, F., Caumon, G., & Renard, P. (2016). Impact of a stochastic sequential  
759 initiation of fractures on the spatial correlations and connectivity of discrete  
760 fracture networks. *Journal of Geophysical Research: Solid Earth*, *121*(8),  
761 5641–5658. doi: 10.1002/2015JB012451
- 762 Bonnet, E., Bour, O., Odling, N. E., Davy, P., Main, I., Cowie, P., & Berkowitz, B.  
763 (2001). Scaling of fracture systems in geological media. *Reviews of geophysics*,  
764 *39*(3), 347–383.
- 765 Bron, C., & Kerbosch, J. (1973). Algorithm 457: finding all cliques of an undirected  
766 graph. *Communications of the ACM*, *16*(9), 575–577. doi: 10.1145/362342  
767 .362367
- 768 Caumon, G. (2010). Towards stochastic time-varying geological modeling. *Mathe-*  
769 *matical Geosciences*, *42*(5), 555–569.
- 770 Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*,  
771 *15*(366), 92–96.
- 772 Cherpeau, N., & Caumon, G. (2015). Stochastic structural modelling in sparse data  
773 situations. *Petroleum Geoscience*, *21*(4), 233–247. doi: 10.1144/petgeo2013  
774 -030
- 775 Cherpeau, N., Caumon, G., Caers, J., & Lévy, B. (2012). Method for stochastic in-  
776 verse modeling of fault geometry and connectivity using flow data. *Mathemati-*  
777 *cal Geosciences*, *44*(2), 147–168. doi: 10.1007/s11004-012-9389-2
- 778 Cherpeau, N., Caumon, G., & Lévy, B. (2010a). Stochastic simulation of fault net-  
779 works from 2D seismic lines. In *SEG Expanded Abstracts* (Vol. 29, pp. 2366–  
780 2370). doi: 10.1190/1.3513325

- 781 Cherpeau, N., Caumon, G., & Lévy, B. (2010b). Stochastic simulations of fault  
782 networks in 3d structural modeling. *Comptes Rendus Géoscience*, *342*(9), 687–  
783 694. doi: 10.1016/j.crte.2010.04.008
- 784 Davy, P., Le Goc, R., & Darcel, C. (2013). A model of fracture nucleation, growth  
785 and arrest, and consequences for fracture density and scaling. *Journal of Geo-*  
786 *physical Research: Solid Earth*, *118*(4), 1393–1407. doi: 10.1002/jgrb.50120
- 787 de la Varga, M., & Wellmann, J. F. (2016, August). Structural geologic modeling  
788 as an inference problem: A Bayesian perspective. *Interpretation*, *4*(3), SM1-  
789 SM16. doi: 10.1190/INT-2015-0188.1
- 790 Edwards, J., Lallier, F., Caumon, G., & Carpentier, C. (2018). Uncertainty  
791 management in stratigraphic well correlation and stratigraphic architec-  
792 tures: A training-based method. *Computers & Geosciences*, *111*, 1–17. doi:  
793 10.1016/j.cageo.2017.10.008
- 794 Emerson. (2018). *SKUA-GOCAD*. [https://www.pdgm.com/products/skua-](https://www.pdgm.com/products/skua-gocad/)  
795 [gocad/](https://www.pdgm.com/products/skua-gocad/).
- 796 Ferrill, D. A., Stamatakos, J. A., & Sims, D. (1999). Normal fault corrugation: Im-  
797 plications for growth and seismicity of active normal faults. *Journal of Struc-*  
798 *tural Geology*, *21*(8), 1027–1038. doi: 10.1016/S0191-8141(99)00017-6
- 799 Frank, T., Tertois, A.-L., & Mallet, J.-L. (2007, July). 3D-reconstruction of complex  
800 geological interfaces from irregularly distributed and noisy point data. *Com-*  
801 *puters & Geosciences*, *33*(7), 932-943. doi: 10.1016/j.cageo.2006.11.014
- 802 Freeman, B., Boulton, P. J., Yielding, G., & Menpes, S. (2010). Using empir-  
803 ical geological rules to reduce structural uncertainty in seismic interpre-  
804 tation of faults. *Journal of Structural Geology*, *32*(11), 1668–1676. doi:  
805 10.1016/j.jsg.2009.11.001
- 806 Freeman, B., Yielding, G., & Badley, M. (1990). Fault correlation during seismic in-  
807 terpretation. *First Break*, *8*(3), 87–95. doi: 10.3997/1365-2397.1990006
- 808 Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical

- 809 science. *Geological Society of America Bulletin*, 107(8), 960–968. doi: 10.1130/  
810 0016-7606(1995)107<0960:GRGAAI>2.3.CO;2
- 811 Gillespie, P., Walsh, J., & Watterson, J. (1992). Limitations of dimension and  
812 displacement data from single faults and the consequences for data analysis  
813 and interpretation. *Journal of Structural Geology*, 14(10), 1157–1172. doi:  
814 10.1016/0191-8141(92)90067-7
- 815 Giraud, J., Lindsay, M., Jessell, M., & Ogarko, V. (2019). Towards geologically rea-  
816 sonable lithological classification from integrated geophysical inverse modelling:  
817 Methodology and application case. *Solid Earth Discussions*, 2019, 1–27. doi:  
818 10.5194/se-2019-164
- 819 Godefroy, G., Caumon, G., Ford, M., Laurent, G., & Jackson, C. A.-L. (2017).  
820 A parametric fault displacement model to introduce kinematic control  
821 into modeling faults from sparse data. *Interpretation*, 6(2), 1–48. doi:  
822 10.1190/int-2017-0059.1
- 823 Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2019). Structural interpre-  
824 tation of sparse fault data using graph theory and geological rules. *Mathemati-  
825 cal Geosciences*.
- 826 Gombert, B., Duputel, Z., Jolivet, R., Doubre, C., Rivera, L., & Simons, M. (2018,  
827 February). Revisiting the 1992 Landers earthquake: A Bayesian exploration  
828 of co-seismic slip and off-fault damage. *Geophysical Journal International*,  
829 212(2), 839–852. doi: 10.1093/gji/ggx455
- 830 Gratier, J.-P., & Guillier, B. (1993, March). Compatibility constraints on folded  
831 and faulted strata and calculation of total displacement using computational  
832 restoration (UNFOLD program). *Journal of Structural Geology*, 15(3-5),  
833 391–402. doi: 10.1016/0191-8141(93)90135-W
- 834 Grose, L., Ailleres, L., Laurent, G., Armit, R., & Jessell, M. (2019). Inversion of ge-  
835 ological knowledge for fold geometry. *Journal of Structural Geology*, 119, 1–14.  
836 doi: 10.1016/j.jsg.2018.11.010

- 837 Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., & Cousin-Dechenaud,  
 838 T. (2018). Inversion of structural geology data for fold geometry. *Journal of*  
 839 *Geophysical Research: Solid Earth*, *123*(8), 6318–6333. doi: doi.org/10.1029/  
 840 2017JB015177
- 841 Henza, A. A., Withjack, M. O., & Schlische, R. W. (2011). How do the properties  
 842 of a pre-existing normal-fault population influence fault development dur-  
 843 ing a subsequent phase of extension? *Journal of Structural Geology*, *33*(9),  
 844 1312–1324. doi: 10.1016/j.jsg.2011.06.010
- 845 Holden, L., Mostad, P., Nielsen, B. F., Gjerde, J., Townsend, C., & Ottesen, S.  
 846 (2003). Stochastic structural modeling. *Mathematical Geology*, *35*(8), 899–914.  
 847 doi: 10.1023/B:MATG.0000011584.51162.69
- 848 Hollund, K., Mostad, P., Nielsen, B. F., Holden, L., Gjerde, J., Contursi, M. G., ...  
 849 Sverdrup, E. (2002). Havana - a fault modeling tool. *Norwegian Petroleum*  
 850 *Society Special Publications*, *11*, 157–171.
- 851 Irakarama, M., Cupillard, P., Caumon, G., Sava, P., & Edwards, J. (2019). Ap-  
 852 praising structural interpretations using seismic data theoretical elements. *Geo-*  
 853 *physics*, *84*(2), N29–N40.
- 854 Irving, A., Chavanne, E., Faure, V., Buffet, P., & Barber, E. (2010). An uncertainty  
 855 modelling workflow for structurally compartmentalized reservoirs. *Geological*  
 856 *Society, London, Special Publications*, *347*(1), 283–299.
- 857 Jessell, M., Aillères, L., & de Kemp, E. A. (2010, July). Towards an integrated in-  
 858 version of geoscientific data: What price of geology? *Tectonophysics*, *490*(3-4),  
 859 294-306. doi: 10.1016/j.tecto.2010.05.020
- 860 Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, J., Hillier, M., ...  
 861 Martin, R. (2014). Next generation three-dimensional geologic modeling and  
 862 inversion. *Economic Geology*, *18*, 261–272.
- 863 Julio, C. (2015). *Conditionnement de la modélisation stochastique 3D des réseaux de*  
 864 *failles* (Unpublished doctoral dissertation). Université de Lorraine.

- 865 Julio, C., Caumon, G., & Ford, M. (2015a). Impact of the en echelon fault connec-  
866 tivity on reservoir flow simulations. *Interpretation*, 3(4), SAC23–SAC34. doi:  
867 10.1190/INT-2015-0060.1
- 868 Julio, C., Caumon, G., & Ford, M. (2015b). Sampling the uncertainty associated  
869 with segmented normal fault interpretation using a stochastic downscaling  
870 method. *Tectonophysics*, 639, 56–67. doi: 10.1016/j.tecto.2014.11.013
- 871 Knuth, D. E. (2005). *The Art of Computer Programming, Volume 4: Generating all*  
872 *Combinations and Partitions, Fascicle 3*. Addison-Wesley Professional.
- 873 Lallier, F., Antoine, C., Charreau, J., Caumon, G., & Ruiu, J. (2013). Manage-  
874 ment of ambiguities in magnetostratigraphic correlation. *Earth and Planetary*  
875 *Science Letters*, 371, 26–36. doi: 10.1016/j.epsl.2013.04.019
- 876 Laurent, G., Caumon, G., Bouziat, A., & Jessell, M. (2013, April). A parametric  
877 method to model 3D displacements around faults with volumetric vector fields.  
878 *Tectonophysics*, 590, 83–93. doi: 10.1016/j.tecto.2013.01.015
- 879 Lecour, M., Cognot, R., Duvinage, I., Thore, P., & Dulac, J.-C. (2001). Mod-  
880 elling of stochastic faults and fault networks in a structural uncertainty study.  
881 *Petroleum Geoscience*, 7(S), S31–S42. doi: 10.1144/petgeo.7.S.S31
- 882 Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions,  
883 and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).
- 884 Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., & Betts, P. G. (2012).  
885 Locating and quantifying geological uncertainty in three-dimensional models:  
886 Analysis of the Gippsland Basin, southeastern Australia. *Tectonophysics*, 546,  
887 10–27. doi: 10.1016/j.tecto.2012.04.007
- 888 Lines, L. R., & Newrick, R. T. (2004). *Fundamentals of Geophysical Interpretation*.  
889 Society of Exploration Geophysicists. doi: 10.1190/1.9781560801726
- 890 Mai, P. M., Galis, M., Thingbaijam, K. K. S., Vyas, J. C., & Dunham, E. M. (2017,  
891 September). Accounting for Fault Roughness in Pseudo-Dynamic Ground-  
892 Motion Simulations. *Pure and Applied Geophysics*, 174(9), 3419–3450. doi:

- 893 10.1007/s00024-017-1536-8
- 894 Mallet, J.-L. (1992). Discrete smooth interpolation in geometric modelling.  
895 *Computer-aided design*, *24*(4), 178–191. doi: 10.1016/0010-4485(92)90054-E
- 896 Manighetti, I., Caulet, C., Barros, L., Perrin, C., Cappa, F., & Gaudemer, Y.  
897 (2015). Generic along-strike segmentation of Afar normal faults, East  
898 Africa: Implications on fault growth and stress heterogeneity on seismogenic  
899 fault planes. *Geochemistry, Geophysics, Geosystems*, *16*(2), 443–467. doi:  
900 10.1002/2014GC005691
- 901 Mann, C. J. (1993). Uncertainty in geology. In J. C. Davis & U. C. E. Herzfeld  
902 (Eds.), *Computers in geology—25 years of progress* (pp. 241–254). Oxford Uni-  
903 versity Press, Inc.
- 904 Manzocchi, T., Heath, A., Childs, C., Telles, I., & Carneiro, M. (2019). Modelling  
905 fault zone displacement partitioning for risking across-fault juxtaposition. In  
906 *81st eage conference and exhibition 2019*.
- 907 Munthe, K., Holden, L., Mostad, P., & Townsend, C. (1994). Modelling sub-seismic  
908 fault patterns using a Marked Point Process. In *Ecmor iv-4th european confer-*  
909 *ence on the mathematics of oil recovery*. doi: 10.3997/2214-4609.201411151
- 910 Nicol, A., Walsh, J., Childs, C., & Manzocchi, T. (2020). The growth of faults. In  
911 *Understanding faults* (pp. 221–255). Elsevier.
- 912 Nixon, C. W., Sanderson, D. J., & Bull, J. M. (2011). Deformation within a strike-  
913 slip fault network at Westward Ho!, Devon UK: Domino vs conjugate faulting.  
914 *Journal of Structural Geology*, *33*(5), 833–843. doi: 10.1016/j.jsg.2011.03.009
- 915 Osypov, K., Yang, Y., Fournier, A., Ivanova, N., Bachrach, R., Yarman, C. E., ...  
916 Woodward, M. (2013). Model-uncertainty quantification in seismic tomogra-  
917 phy: Method and applications. *Geophysical Prospecting*, *61*(6), 1114–1134.
- 918 Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., & Jessell, M. (2018,  
919 April). Monte Carlo simulation for uncertainty estimation on structural  
920 data in implicit 3-D geological modeling, a guide for disturbance distri-



- 921           bution selection and parameterization.     *Solid Earth*, 9(2), 385–402.     doi:  
922           10.5194/se-9-385-2018
- 923     Pakyuz-Charrier, E., Jessell, M., Giraud, J., Lindsay, M., & Ogarko, V.     (2019).  
924           Topological analysis in monte carlo simulation for uncertainty estimation. *Solid*  
925           *Earth Discussions*, 2019, 1–37. doi: 10.5194/se-2019-78
- 926     Peacock, D., & Sanderson, D.     (1991).     Displacements, segment linkage and relay  
927           ramps in normal fault zones.     *Journal of Structural Geology*, 13(6), 721–733.  
928           doi: 10.1016/0191-8141(91)90033-F
- 929     Ragon, T., Sladen, A., & Simons, M.     (2018).     Accounting for uncertain fault geom-  
930           etry in earthquake source inversions-i: theory and simplified application. *Geo-*  
931           *physical Journal International*, 214(2), 1174–1190.
- 932     Richards, F. L., Richardson, N. J., Bond, C. E., & Cowgill, M.     (2015).     Interpre-  
933           tational variability of structural traps: implications for exploration risk and  
934           volume uncertainty.     *Geological Society, London, Special Publications*, 421(1),  
935           7–27. doi: 10.1144/SP421.13
- 936     Riesner, M., Durand-Riard, P., Hubbard, J., Plesch, A., & Shaw, J. H.     (2017, May).  
937           Building Objective 3D Fault Representations in Active Tectonic Settings. *Seis-*  
938           *mological Research Letters*, 88(3), 831-839. doi: 10.1785/0220160192
- 939     Rivenæs, J. C., Otterlei, C., Zachariassen, E., Dart, C., & Sjøholm, J.     (2005).     A 3D  
940           stochastic model integrating depth, fault and property uncertainty for plan-  
941           ning robust wells, Njord Field, offshore Norway.     *Petroleum Geoscience*, 11(1),  
942           57–65. doi: 10.1144/1354-079303-612
- 943     Røe, P., Georgsen, F., & Abrahamsen, P.     (2014).     An uncertainty model for fault  
944           shape and location. *Mathematical Geosciences*, 46(8), 957–969.
- 945     Rosenbaum, M. S., & Culshaw, M. G.     (2003).     Communicating the risks arising from  
946           geohazards. *Journal of the Royal Statistical Society: Series A (Statistics in So-*  
947           *cety)*, 166(2), 261–270. doi: 10.1111/1467-985X.00275
- 948     Rotevatn, A., Jackson, C. A.-L., Tvedt, A. B., Bell, R. E., & Blækkan, I.     (2018, Au-

- 949           gust). How do normal faults grow? *Journal of Structural Geology*. doi: 10  
950           .1016/j.jsg.2018.08.005
- 951 Sanfeliu, A., & Fu, K.-S. (1983). A distance measure between attributed relational  
952           graphs for pattern recognition. *IEEE transactions on systems, man, and cyber-*  
953           *netics*(3), 353–362.
- 954 Schaaf, A., & Bond, C. E. (2019). Quantification of uncertainty in 3-d seismic in-  
955           terpretation: implications for deterministic and stochastic geomodelling and  
956           machine learning. *Solid earth*.
- 957 Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, 1(1), 27–64. doi:  
958           10.1016/j.cosrev.2007.05.001
- 959 Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface sys-*  
960           *tems* (Vol. 236). John Wiley & Sons.
- 961 Schneeberger, R., de La Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F.,  
962           & Herwegh, M. (2017). Methods and uncertainty estimations of 3-d struc-  
963           tural modelling in crystalline rocks: a case study. *Solid Earth*, 8(5), 987. doi:  
964           10.5194/se-8-987-2017
- 965 Seiler, A., Aanonsen, S. I., Evensen, G., & Rivenæs, J. C. (2010). Structural sur-  
966           face uncertainty modeling and updating using the ensemble Kalman filter. *SPE*  
967           *Journal*, 15(04), 1–062. doi: 10.2118/125352-PA
- 968 Sepúlveda, I., Liu, P. L.-F., Grigoriu, M., & Pritchard, M. (2017, September).  
969           Tsunami hazard assessments with consideration of uncertain earthquake slip  
970           distribution and location: TSUNAMI HAZARD AND UNCERTAIN EARTH-  
971           QUAKES. *Journal of Geophysical Research: Solid Earth*, 122(9), 7252–7271.  
972           doi: 10.1002/2017JB014430
- 973 Smith, T. F., & Waterman, M. S. (1980). New stratigraphic correlation techniques.  
974           *The Journal of Geology*, 88(4), 451–457.
- 975 Stoyan, D., & Gloaguen, R. (2011, August). Nucleation and growth of geological  
976           faults. *Nonlinear Processes in Geophysics*, 18(4), 529–536. doi: 10.5194/npg-18

977 -529-2011

- 978 Suzuki, S., Caumon, G., & Caers, J. (2008). Dynamic data integration for struc-  
 979 tural modeling: model screening approach using a distance-based model  
 980 parameterization. *Computational Geosciences*, *12*(1), 105–119. doi:  
 981 10.1007/s10596-007-9063-9
- 982 Tal, Y., Hager, B. H., & Ampuero, J. P. (2018, January). The Effects of Fault  
 983 Roughness on the Earthquake Nucleation Process. *Journal of Geophysical  
 984 Research: Solid Earth*, *123*(1), 437-456. doi: 10.1002/2017JB014746
- 985 Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, *2*(8),  
 986 492–494. doi: 10.1038/nphys375
- 987 Thibaut, M., Gratier, J. P., Léger, M., & Morvan, J. M. (1996). An inverse method  
 988 for determining three-dimensional fault geometry with thread criterion: Appli-  
 989 cation to strike-slip and thrust faults (Western Alps and California). *Journal  
 990 of Structural Geology*, *18*(9), 1127-1138.
- 991 Thiele, S. T., Jessell, M. W., Lindsay, M., Wellmann, J. F., & Pakyuz-Charrier, E.  
 992 (2016, October). The topology of geology 2: Topological uncertainty. *Journal  
 993 of Structural Geology*, *91*, 74-87. doi: 10.1016/j.jsg.2016.08.010
- 994 Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T., & Cognot, R. (2002). Struc-  
 995 tural uncertainties: Determination, management, and applications. *Geophysics*,  
 996 *67*(3), 840–852. doi: 10.1190/1.1484528
- 997 Torabi, A., & Berg, S. S. (2011). Scaling of fault attributes: A review. *Marine and  
 998 Petroleum Geology*, *28*(8), 1444–1460. doi: 10.1016/j.marpetgeo.2011.04.003
- 999 Tvedt, A. B., Rotevatn, A., & Jackson, C. A. (2016). Supra-salt normal fault  
 1000 growth during the rise and fall of a diapir: Perspectives from 3D seismic reflec-  
 1001 tion data, Norwegian North Sea. *Journal of Structural Geology*, *91*, 1–26. doi:  
 1002 10.1016/j.jsg.2016.08.001
- 1003 Wellmann, J. F., & Caumon, G. (2018, Nov). 3-D structural geological models:  
 1004 Concepts, methods, and uncertainties. In I. S. C. (Ed) (Ed.), *Advances in geo-*

- 1005 *physics* (Vol. 59, p. 1-121). Elsevier. doi: 10.1016/bs.agph.2018.09.001
- 1006 Wellmann, J. F., Horowitz, F. G., Schill, E., & Regenauer-Lieb, K. (2010, July). To-  
 1007 wards incorporating uncertainty of structural data in 3D geological inversion.  
 1008 *Tectonophysics*, 490(3-4), 141–151. doi: 10.1016/j.tecto.2010.04.022
- 1009 Wellmann, J. F., Lindsay, M., Poh, J., & Jessell, M. (2014). Validating 3-D struc-  
 1010 tural models with geological knowledge for improved uncertainty evaluations.  
 1011 *Energy Procedia*, 59, 374–381. doi: 10.1016/j.egypro.2014.10.391
- 1012 Wilson, C. G., Bond, C. E., & Shipley, T. F. (2019). How can geologic decision mak-  
 1013 ing under uncertainty be improved? *Solid earth*, 1–34.
- 1014 Wood, D. R. (2007, June). On the Maximum Number of Cliques in a Graph. *Graphs*  
 1015 *and Combinatorics*, 23(3), 337-352. doi: 10.1007/s00373-007-0738-8
- 1016 Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., & Caumon, G. (2019, Octo-  
 1017 ber). Building realistic structure models to train convolutional neural net-  
 1018 works for seismic structural interpretation. *GEOPHYSICS*, 1-48. doi:  
 1019 10.1190/geo2019-0375.1
- 1020 Yielding, G. (2016). The geometry of branch lines. *Geological Society, London, Spe-*  
 1021 *cial Publications*, 439, SP439–1. doi: 10.1144/SP439.1
- 1022 Zakian, P., Khaji, N., & Soltani, M. (2017, October). A Monte Carlo adapted  
 1023 finite element method for dislocation simulation of faults with uncertain  
 1024 geometry. *Journal of Earth System Science*, 126(7), 105. doi: 10.1007/  
 1025 s12040-017-0878-z
- 1026 Zehner, B., Börner, J. H., Görz, I., & Spitzer, K. (2015, June). Workflows  
 1027 for generating tetrahedral meshes for finite element simulations on com-  
 1028 plex geological structures. *Computers & Geosciences*, 79, 105-117. doi:  
 1029 10.1016/j.cageo.2015.02.009
- 1030 Zhu, S., Hack, R., Turner, K., & Hale, M. (2003). How far will uncertainty of the  
 1031 subsurface limit the sustainability planning of the subsurface? In *Proc. sus-*  
 1032 *tainable development & management of the subsurface (sdms) conference* (pp.

