



**HAL**  
open science

# From fault likelihood to fault networks: stochastic seismic interpretation through a marked point process with interactions

F. Taty Moukati, Radu Stefan Stoica, François Bonneau, Xinming Wu,  
Guillaume Caumon

## ► To cite this version:

F. Taty Moukati, Radu Stefan Stoica, François Bonneau, Xinming Wu, Guillaume Caumon. From fault likelihood to fault networks: stochastic seismic interpretation through a marked point process with interactions. 2024. hal-04513328

**HAL Id: hal-04513328**

<https://hal.univ-lorraine.fr/hal-04513328v1>

Preprint submitted on 20 Mar 2024

**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.

# From fault likelihood to fault networks: stochastic seismic interpretation through a marked point process with interactions

Fabrice Taty Moukati · Radu Stefan Stoica · François Bonneau · Xinming Wu · Guillaume Caumon

February 5, 2024

**Abstract** Faults are crucial subsurface features that significantly influence the mechanical behavior and hydraulic properties of rock masses. Interpreting them from seismic data may lead to various scenarios due to uncertainties arising from limited seismic bandwidth and possible imaging errors. Although methods addressing fault uncertainty exist, only a few of them can produce curved and sub-seismic faults at once while quantitatively honoring seismic images and avoiding anchoring in a reference interpretation. In this work, we use a mathematical framework of marked point processes to approximate fault networks in two dimensions with a set of line segments. The proposed stochastic model, namely Candy Model, incorporates simple pairwise and nearby connections to capture the interactions between fault segments. The novelty of this approach lies in conditioning the stochastic model using input image of fault probabilities generated by a Convolutional Neural Network (CNN). The Metropolis-Hastings algorithm is used to generate various scenarios of fault network configurations, thereby exploring the model space associated with the Candy Model and reflecting the uncertainty. Probability level sets constructed from these fault segment configurations provide insights on the obtained realizations and on the model parameters. The empty space function pro-

---

Fabrice Taty Moukati  
GeoRessources - ENSG, Université de Lorraine - CNRS, F-54000 Nancy  
2 rue du Doyen Marcel Roubault. F-54518 VANDOEUVRE-LES-NANCY  
E-mail: tatyfabrice@gmail.com

Radu Stefan Stoica  
IECL, Université de Lorraine - CNRS - Inria, F-54500 Nancy France  
B.P. 70239, 54506 VANDOEUVRE-LES-NANCY, CEDEX

François Bonneau  
GeoRessources - ENSG, Université de Lorraine - CNRS, F-54000 Nancy  
2 rue du Doyen Marcel Roubault. F-54518 VANDOEUVRE-LES-NANCY

Xinming Wu  
University of Science and Technology of China, Hefei

Guillaume Caumon  
GeoRessources - ENSG, Université de Lorraine - CNRS, F-54000 Nancy  
2 rue du Doyen Marcel Roubault. F-54518 VANDOEUVRE-LES-NANCY  
Institut Universitaire de France (IUF)

duces a ranking of the generated fault networks against an existing interpretation by testing and quantifying their spatial variability. The approach is applied on two-dimensional sections of seismic data, acquired in the Central North Sea.

**Keywords** Uncertainty · Seismic interpretation · Statistics · Image analysis · Faults · Machine learning

## 1 Introduction

Faults are mechanical discontinuities which impact our understanding of the history of the Earth, the mechanical properties of rock masses, the hydrodynamic behavior of reservoirs and aquifers, and the nucleation and rupture of earthquakes. This makes the forecasting of fault location and fault architecture very important in mining, geotechnical projects (Islam and Shinjo, 2009), reservoir engineering (Manzocchi et al., 1999), hydrogeology (Bense et al., 2013) and earthquake science (Rosenbaum and Culshaw, 2003; Gabriel et al., 2013).

Reflection seismic images are widely used to map faults in the subsurface. Their quality depends on wave frequency (Botter et al., 2017; Chopra et al., 2006) and is sensitive to migration parameters (Faleide et al., 2021a; Thore et al., 2002). Detecting faults with a small vertical displacement calls for a good vertical resolution (Dimmen et al., 2023), which is crucial for evaluating 3D fault architecture (Botter et al., 2017; Julio et al., 2015). An additional difficulty is that faults are often segmented (i.e. composed of multiple almost aligned subplanar cracks, Childs et al., 2009a), and the spacing between segments may be below the scale accessible with seismic imaging. These physical limitations imply that deterministic fault interpretations from seismic images may underestimate topological fault uncertainty.

Faults and horizons are usually mapped by extracting inline and crossline sections from a 3D seismic cube, with varying increment depending on the continuity, strength, structural complexity of the reflections (Faleide et al., 2021b). Faults are then detected from breaks in reflection continuity (Dimmen et al., 2023). Manual interpretation can be tedious and yields a unique solution, which may be biased (Bond, 2015; Froner et al., 2013) since it relies on visual perception and experience built up over years (Froner et al., 2013). To avoid possible subjective biases and reduce interpretation time, automatic seismic interpretation and fault attribute volumes (e.g., Bedle and van Gestel, 2021; Wu et al., 2020; Iacopini and Butler, 2011; Manzi et al., 2020) have been proposed. However, most automatic interpretation methods are pixel-based, so are prone to become less effective when the contextual information in the close surroundings of sought objects varies a lot (Descombes, 2012).

Complementary to deterministic fault interpretation, stochastic discrete modeling approaches have been used both in the context of seismic interpretation (Julio et al., 2015; Aydin and Caers, 2017; Holden et al., 2003; Goodwin et al., 2022) and in the context of sparse fault observations (Godefroy et al., 2021; Soleng et al., 2004; Cherpeau et al., 2010; Cherpeau and Caumon, 2015) to generate several fault interpretation scenarios reflecting uncertainty. Most of these methods are based on marked point processes (MPPs) to simulate faults (e.g., Strauss process, see Aydin and Caers, 2017). In this work, we tackle fault segmentation while

quantifying the uncertainty associated to the number and the connectivity between fault segments honoring a given seismic dataset.

Using marked point processes to represent faults can raise challenges due to the difficulty of capturing the interactions between fault objects (e.g., [Bonneau et al., 2016](#)). Poisson point processes in which each fault is independently generated from each other may well reflect the initiation of faults but fail to integrate the local interactions (characteristics, geometry and behaviors at a more detailed level) that may exist between fault objects. Strauss point processes, which imposes repulsions between marked points ([Aydin and Caers, 2017](#)) and parent-daughter processes ([Bonneau et al., 2016](#)) have been proposed to better integrate these interactions from a statistical point of view. However, methods characterizing faults while quantitatively honoring seismic images have not received much attention. An original curved support marked point process has been proposed to approximate faults with Bezier curves in vertical sections of seismic data ([Barna et al., 2015](#)). This model, given some seismic attribute volumes, is able to simultaneously model both sub-seismic and curved faults. Whereas the definition of a data term based on seismic attributes is suitable for both fault detection and uncertainty quantification, the exact details of repulsions between Bezier curves, remains unclear and unlikely to capture the detailed behaviors and orientations of sub-scale fault features. In our approach, we use a simpler representation of straight line segments to approximate faults based on some hypotheses: (1) major or large faults are segmented into smaller and interconnected segments (e.g., [Roche et al., 2021](#)), and (2) the fault segments are almost aligned ([Childs et al., 2009b](#)). On that basis, we propose to use the connection and alignment interactions of the Candy Model ([Stoica et al., 2005](#)) to obtain configurations of segments exhibiting a topology close to that of observed faults (Section 3.1). Fault localization is governed by specific data conditioning strategies in the MPP framework. Whereas previous works have considered expert-based interpretations ([Cherpeau and Caumon, 2015](#); [Aydin and Caers, 2017](#)) in 3D, [Barna et al. \(2015\)](#) consider fault attributes such as coherence and gradient disorder. As for [Goodwin et al. \(2022\)](#), displacement and strike orientations are used to guide the simulation of faults. In this paper, we use a fault likelihood as computed by the Convolutional Neural Network (CNN) FaultNet3D ([Wu et al., 2019](#)), and for which we develop a data term for the Candy Model (Section 3.4). FaultNet3D model computes the probability and orientation of faults at every single location in a given seismic image (Section 2). To implement our method, we use the Metropolis-Hastings (MH) algorithm tailored to the Candy Model (Section 3.5), to sample various possible fault network configurations.

In Section 4, we apply the proposed approach to the Volve seismic data, acquired in the central part of the North Sea ([Tunkiel et al., 2020](#); [Ravasi et al., 2015](#); [Szydluk et al., 2006](#)), which is 5 km north of Sleipner East field. Before applying the Candy Model to the Volve seismic data, a simulation study is performed to analyze the stationarity of the Markov Chain built by the Metropolis-Hastings algorithm tailored to this model (Section 4.1). To capture the fault uncertainty and explore the model parameters space associated with the Candy Model, various possible fault network realizations (Section 4.2) are generated by the MH algorithm. To communicate and visualize this uncertainty, the level sets (or visit map) of these fault network realizations are constructed. Finally, we compare the obtained results with the skeleton extracted from the fault likelihood ([Wu et al., 2019](#)) by testing and quantifying the spatial distribution of areas without fault

objects (Section 4.3), using a summary statistic named the empty space function  $F$  or the spherical contact distribution (Kleinschroth et al., 2017).

## 2 Input data

### 2.1 Fault likelihood

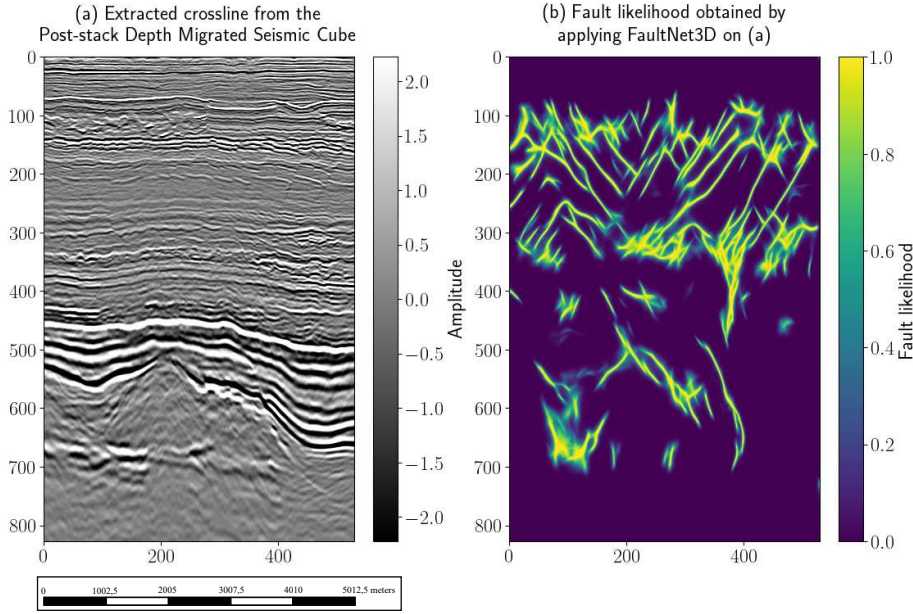
The large diversity of faults in the subsurface makes their manual and automatic interpretation error-prone and time-consuming. The key issue is that manual interpretation is unique and difficult to reproduce. This has motivated specific attributes to help interpreters to detect faults or to automate the process. Methods to compute fault likelihood may be based on several means of detecting seismic discontinuities such as amplitude-based, semblance-based or neural network-based methods (Marfurt et al., 1998; Ashraf et al., 2020; Imran et al., 2021; Hale, 2013; Jacquemin and Mallet, 2005; Wu et al., 2019). The semblance-based fault attribute is computed by scanning the semblance volume along multiple fault orientations and retaining the best alignment. This process produces images of the maximum fault likelihood and the corresponding strike and dip angles. In FaultNet3D (Wu et al., 2019), the CNN-based fault likelihood is performed by firstly predicting the dip and strike orientation angles automatically at every voxel of the seismic image by a CNN trained on small synthetic faulted models. Then, these predicted fault orientations are aggregated to produce the fault likelihood cube by local convolutions. The fault likelihood attribute is regarded as the probability that a fault is present at every location of the seismic data.

### 2.2 Fault likelihood application

In this work, we use the Volve poststack depth-migrated seismic cube provided by Equinor, acquired, in 2002, in the gas-/condensate-rich Sleipner area of the Central North Sea (STATOIL, 1993; Das et al., 2019; Tunkiel et al., 2020; Ravasi et al., 2015; Szydlik et al., 2006). Volve is a small dome-shaped structure formed by the collapse of adjacent salt ridges during Jurassic (Ravasi et al., 2015; Szydlik et al., 2006). In this paper, we focus on the fault network in the overlying Cretaceous and Cenozoic units. We ran the FaultNet3D model on these data to produce a fault probability image (Fig. 1b). Fig. 1b shows clear and continuous fault-like shapes to identify fault positions compared to the seismic amplitude image (Fig. 1a). More sections and the use of this fault likelihood in the proposed stochastic model will be presented in Section 4.

## 3 Methodology

In this section, we present the proposed mathematical formalism to interpret fault networks from the fault likelihood. First of all, we give an overview of marked point processes (Section 3.1), inspired by the detailed presentations of Stoica et al. (2005), van Lieshout and Stoica (2003), Stoica et al. (2007), Tempel et al. (2014), Chiu et al. (2013), and Stoica (2014). In Section 3.2, we give the description of



**Fig. 1** (a) Volve seismic data set provided by Equinor given as an input to FaultNet3D, and the corresponding output (b) fault probabilities.

the Candy Model used as a prior to characterize and describe fault geometry and relationships between fault segments. Section 3.3 shows how to choose the model parameters. Finally, Section 3.4 presents the data term to honor the fault likelihood, and Section 3.5 describes a sampler able to simulate the model.

### 3.1 Marked point processes: some theoretical tools

Point processes are probabilistic models for random configurations of points. If marks are attached to the points, we speak about marked point processes. If the marks represent geometrical characteristics of an object, the process is also called an object point process. This modeling strategy considers the faults mapped from seismic images as a realization of a segment point process.

Let  $W$  be a planar observation window of area  $\nu(W) \in [0, +\infty[$ . A segment is represented as a point  $c \in W$  (which is the segment center) with an attached mark taking values in the Cartesian product  $M = [l_{\min}, l_{\max}] \times [0^\circ, 180^\circ[$  for some  $0 < l_{\min} < l_{\max} < +\infty$ , with its orientation  $a \in [0^\circ, 180^\circ[$ , and its length  $l \in [l_{\min}, l_{\max}]$ . A fault network is therefore a finite set  $\mathbf{x} = \{(c_1, (a_1, l_1)), (c_2, (a_2, l_2)), \dots, (c_n, (a_n, l_n)), i = 1, \dots, n\}$  of  $n \in \mathbb{N}$  segments  $s_i \in S = W \times M$ . The model space can then be written as  $\Omega = \cup_{n=0}^{+\infty} S_n$ , with  $S_n$  the set of all (unordered) configurations of segments  $\mathbf{x}$  that consists of  $n$ , not necessarily distinct, segments  $s_i \in S$ .

The best known example of marked point processes is the unit rate Poisson process on  $W \times M$ , with the locations distributed according to a Poisson process with intensity  $\nu(W)$  and independent and identically distributed marks according

to a certain distribution  $\nu_M$ . The Poisson process is used as a reference to generalize more complex processes with interactions between points or objects, which are constructed by specifying the probability density

$$p(\mathbf{x}|\theta) = \frac{\exp[-U(\mathbf{x}|\theta)]}{Z(\theta)}, \quad (1)$$

where  $\theta$  denotes the model parameters,  $Z(\theta)$  the normalizing constant and  $U(\mathbf{x}|\theta)$  the interaction energy of the system. This energy function is written as the sum

$$U(\mathbf{x}|\theta) = \sum_{j=1}^q \sum_{\{s_1, \dots, s_j\} \subseteq \mathbf{x}} \omega^{(j)}(s_1, \dots, s_j), \quad (2)$$

where  $\omega^{(j)} : (W \times M)^j \rightarrow \mathbb{R}$  for  $j = 1, \dots, q$  are the interaction potentials ( $q$  represents the maximum order of all interactions between objects). The marked point process described by (1) is known in the literature as Gibbs point process (Chiu et al., 2013; Baddeley et al., 2016). If there exists a positive real  $C > 0$  such that  $U(\mathbf{x}|\theta) - U(\mathbf{x} \cup \{(c, (a, l))\}|\theta) \leq \log C$  for all  $(c, (a, l)) \in W \times M$ , the process is said to be locally stable. This implies the Ruelle stability condition which ensures the integrability of (1). This local stability is essential in establishing convergence proofs for the Monte Carlo algorithms simulating such a model.

The fault likelihood (Fig. 1) displays probabilities of fault presence, in which every single location of non-zero probability may host a segment center. Our hypothesis is that the connection and the aligned interactions between segments of the Candy Model (Stoica et al., 2005) may exhibit a configuration of segments that has a topology close to the one of the underlying fault network imaged by the fault likelihood. Based on this assumption, the energy function of Equation (1) writes

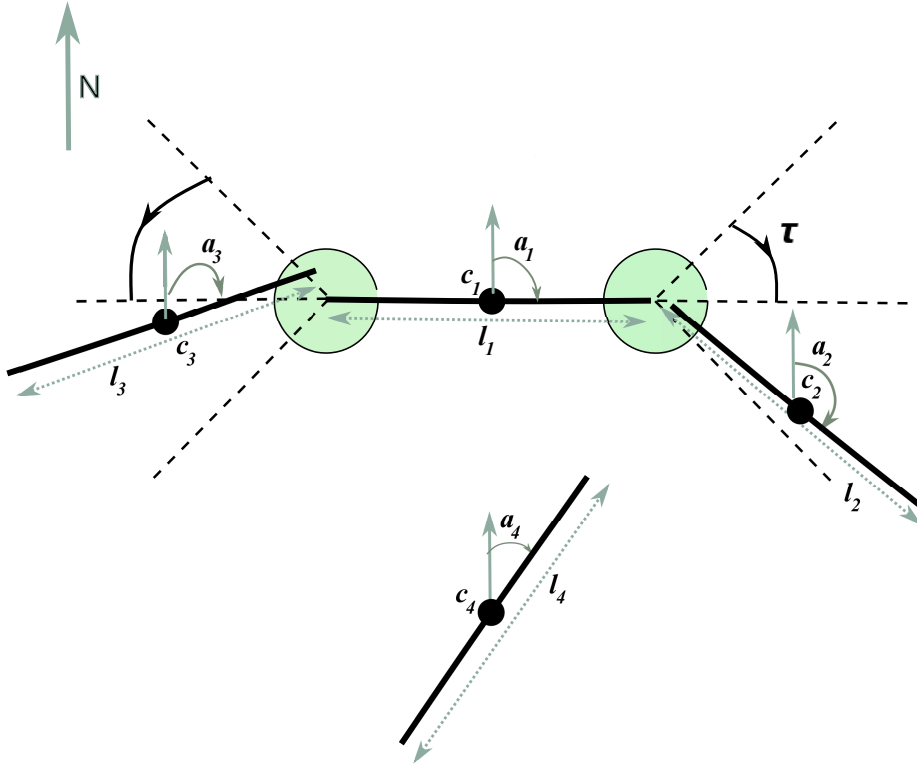
$$p(\mathbf{x}|\theta) \propto \exp[-(U_d(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta))], \quad (3)$$

where  $U_d(\mathbf{x}|\theta)$  and  $U_i(\mathbf{x}|\theta)$  are respectively the data energy and the interaction energy.

### 3.2 Interaction energy : the Candy Model

The interaction term is defined by the weighted sum of terms of the Candy Model (Stoica, 2014; Stoica et al., 2021; van Lieshout and Stoica, 2003; Stoica et al., 2005, 2001, 2004), which measure the alignments and the connections of segments forming the fault network. These interactions are based on geometric parameters.

*Connection interaction.* The circles of radius  $r_c$  centered on both extremities of a segment define the connection region of the segment (Fig. 2): Two segments  $s_i$  and  $s_j$  are connected ( $s_i \sim_c s_j$ ) if only one extremity of a segment is in the connection region of the other segment and if  $\|a_i - a_j\| \leq \tau$ . The parameter  $\tau$  controls the maximum orientation difference of two connected segments. It is called the maximum curvature threshold. This interaction yields three possible types of segments : free segments (e.g.,  $s_4$  in Fig. 2) which are isolated, singly connected segments (e.g.,  $s_2$  in Fig. 2) and double connected segments (e.g.,  $s_1$  in Fig. 2).



**Fig. 2** Connection interaction of the Candy Model.  $s_1 \sim_c s_2$  and  $s_1 \sim_c s_3$ , so  $s_1$  is a doubly connected segment,  $s_4$  is free and  $s_2$  is singly connected.

Patterns of connected and almost aligned objects are produced by privileging doubly connected segments while penalizing both free and singly connected segments. However, this type of interaction does not suffice to have patterns of connected segments and structures as segments may be very close or intersect. The shadowing interaction below prevents some intersections of two close segments from occurring (Stoica et al., 2001).

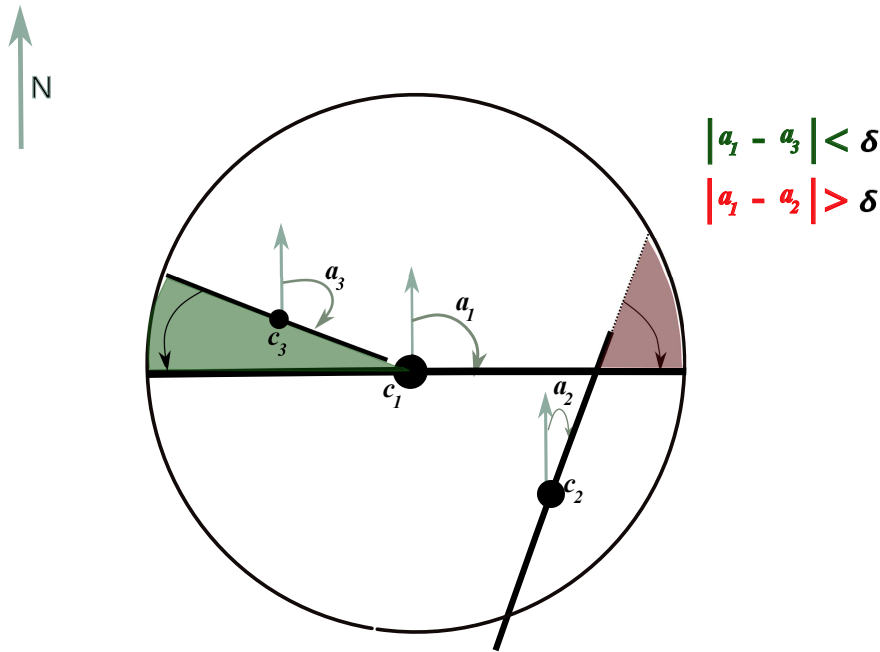
*Shadowing interaction.* For a segment  $(c, l, a)$ , the circle centered on  $c$  and of radius  $l/2$  represents the shadowing region. Two segments  $s_i$  and  $s_j$  are considered in shadowing interaction ( $s_i \sim_r s_j$ ) if and only if

$$d(c_i, c_j) < \max\{l_i, l_j\}/2 \text{ and } \min\{|a_i - a_j|, 180^\circ - |a_i - a_j|\} < \delta, \quad (4)$$

where  $c_i$  and  $c_j$  are the centers of the segments  $s_i$  and  $s_j$ , respectively.  $d(c_i, c_j)$  is the Euclidean distance between the segment centers. For instance, segments  $s_1$  and  $s_3$  are in shadowing interaction and segments  $s_1$  and  $s_2$  are not in shadowing interaction (Fig. 3). The configurations containing segments in shadowing interaction are strongly penalized by the model.

The number of each type of interacting segments is controlled by model parameters. The quantities  $n_f(\mathbf{x})$ ,  $n_s(\mathbf{x})$ ,  $n_d(\mathbf{x})$ ,  $n_r(\mathbf{x})$  denote the numbers of free, singly





**Fig. 3** Shadowing interaction of the Candy Model. The hypothesis of the shadow interaction on the distance between  $s_1$  and  $s_2$ , and between  $s_1$  and  $s_3$  is satisfied. However, only the segments  $s_1$  and  $s_3$  form a configuration of rejected segments (modified from Lacoste, 2004).

connected, doubly connected segments, and of segments in shadowing interaction, respectively. The interaction energy to use in Equation (3) writes

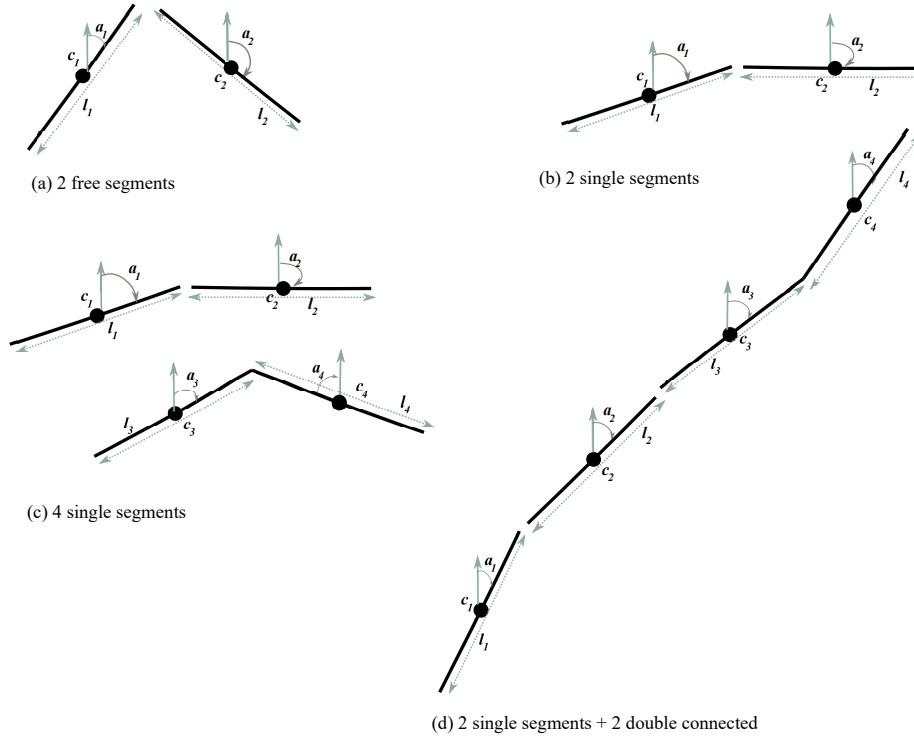
$$U_i(\mathbf{x}|\theta) = -n_d(\mathbf{x}) \log \gamma_d - n_f(\mathbf{x}) \log \gamma_f - n_s(\mathbf{x}) \log \gamma_s - n_r(\mathbf{x}) \log \gamma_r, \quad (5)$$

where  $\mathbf{x}$  is a configuration of segments. Equivalently, the Gibbsian probability distribution of the model is written

$$p_i(\mathbf{x}|\theta) \propto \exp[-U_i(\mathbf{x}|\theta)] = \gamma_d^{n_d(\mathbf{x})} \gamma_f^{n_f(\mathbf{x})} \gamma_s^{n_s(\mathbf{x})} \gamma_r^{n_r(\mathbf{x})}. \quad (6)$$

### 3.3 Choosing the Candy Model parameters

In the Candy Model, two main parameter sets can be distinguished (Lacoste et al., 2004): one refers to physical parameters (e.g., segment length, see more details in section 4.2.2), and the second refers to the interaction parameters in equations (5-6). Several methods exist to estimate them but they demand the knowledge of the pattern. Here, we only focus on the Qualitative Boxes method (Azencott, 1988; Azencott et al., 1995) that has been used by Stoica (2001); Lacoste (2004) to calibrate these parameters. The “qualitative box” for the interaction parameters is a technique for rough guesses allowing a fast evaluation of energy functions (Azencott, 1988) and providing linear inequalities of such parameters (Azencott et al., 1995).



**Fig. 4** Model parameter estimation. (a) Configuration of two free segments, noted as  $\mathbf{x}_a$ , (b) configuration of two singly segments, noted as  $\mathbf{x}_b$ , (c) configuration of two singly and one free segments, noted as  $\mathbf{x}_c$ , and (d) configuration of two singly and one double connected segments, noted as  $\mathbf{x}_d$ . Green arrow represents the North.

Fig. 4 shows different fault configurations for which one can choose the model parameters based on whether or not some configurations are privileged as compared to others. The energy function (5) enables us to formalize this idea.

1. If one wishes to favor configuration with few singly connected segments but more than isolated ones, the energy of a configuration of two isolated segments should be greater than that of a configuration containing two singly connected ones:  $U(\mathbf{x}_a) > U(\mathbf{x}_b) \Leftrightarrow -2 \log \beta - 2 \log \gamma_f > 2 \log \beta - 2 \log \gamma_s \Leftrightarrow \log \gamma_f < \log \gamma_s$ .
2. Setting  $\gamma_r = 0$  entails no shadowing interacting segments.
3. If one wishes more doubly connected and aligned fault segments, the energy of a configuration of four singly connected segments should be greater than that of a configuration containing two singly and two doubly connected fault segments:  $U(\mathbf{x}_c) > U(\mathbf{x}_d) \Leftrightarrow -4 \log \beta - 4 \log \gamma_s > -4 \log \beta - 2 \log \gamma_s - 2 \log \gamma_d \Leftrightarrow 2 \log \gamma_s < \log \gamma_d$ .

Meeting the above requirements yields  $\gamma_f < \gamma_s < \gamma_d$ , with  $\gamma_f, \gamma_s, \gamma_d \in [0, +\infty[$ .

### 3.4 Data energy

Classical applications of the MPP framework use the external-field approach (Descombes, 2012; Lacoste, 2004; Stoica et al., 2001; Barna et al., 2015), measuring the contrast between the background image and the sought objects. In this case study, we use the output of a CNN-based model, similar to Mabon et al. (2023). The data term  $U_d(\mathbf{x})$  is related to the positions of segments on the fault likelihood image. It is based on the rasterization of segments on the fault likelihood image, which is used to compute the segment's potential based on segment data (likelihood) scores. To compute the segment data score, a segment  $s$  is rasterized (Fig. 5) on the likelihood image. This is obtained by the Bresenham algorithm (Bresenham, 1965). The algorithm computes the pixels of the image intersected by the segment, which is represented by the application

$$\begin{aligned} \phi : W \times M &\rightarrow \mathbb{Z}^{2 \times q} \\ s = (c, l, a) &\mapsto \phi(s) = \{(u_i, v_i), i = \{1, \dots, q\}\}, \end{aligned} \quad (7)$$

where  $q$  is the number of pixels touched by the segment  $s$  and  $(u_i, v_i)$  the pixel coordinates of the  $i^{\text{th}}$  pixel. Let the application

$$\begin{aligned} \psi : \mathbb{Z}^{2 \times q} &\rightarrow \mathbb{R}^q \\ \phi(s) &\mapsto \psi(\phi(s)) = \{\psi(u_i, v_i) = \psi_i, i = \{1, \dots, q\}\} \end{aligned}$$

give the probabilities for the pixels to intersect a given segment  $s$ . From this discrete representation, the idea is to get the lowest fault probability value of the pixels traversed by the candidate segment and check if this value is higher than a given fault likelihood threshold  $\sigma$ . This threshold makes it possible to make the simulation of faults more selective by filtering fault likelihoods deemed too low. In the context of uncertainty assessment, this value should remain close to zero to avoid underestimating the fault length. The candidate segment is filtered as follows

$$\mathbb{1}\{\min_{1 \leq i \leq q} (\psi_i)_q \geq \sigma\} = \prod_{i=1}^q \mathbb{1}\{\psi_i \geq \sigma\}. \quad (8)$$

Checking if the fault probability values  $\psi_i$  are all greater than  $\sigma$  is equivalent to checking if the minimum of those values is greater than  $\sigma$ . Finally, we define the segment data score  $v(\phi(s))$  corresponding to the mean fault likelihood of a segment  $s$  as

$$v(\phi(s)) = \mathbb{1}\{\min_{1 \leq i \leq q} (\psi_i)_q \geq \sigma\} \frac{1}{q} \sum_{i=1}^q \psi_i. \quad (9)$$

The total data energy

$$U_d(\mathbf{x}) = \lambda_d \sum_{s \in \mathbf{x}} V_d(\{s\}) \quad (10)$$

is then obtained by scanning the data energy contribution  $V_d(\{s\})$  of each segment in the configuration

$$V_d(\{s\}) = \begin{cases} V_{\max} & \text{if } v(\phi(s)) = 0 \\ -v(\phi(s)) & \text{if not} \end{cases} \quad (11)$$

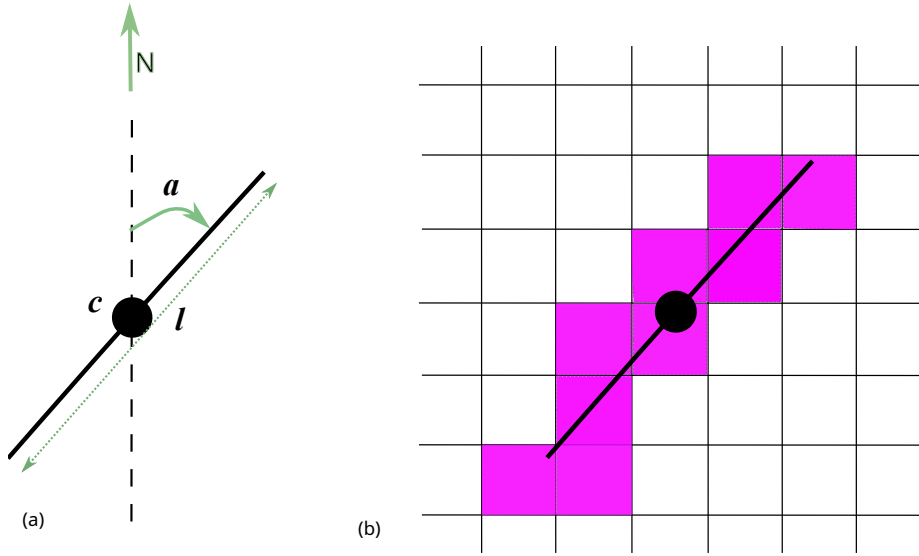


Fig. 5 Representation of a segment in two dimensions.

with  $V_{\max} > 0$  is a fixed large energy value. This total data energy combines two ideas. One is related to the scaling factor  $\lambda_d$  used to balance the *a priori* model energy with respect to the data energy (Lafarge and Descombes, 2012). The second idea is to use a fixed large energy value for a segment lying in a region where the likelihood is null (Stoica et al., 2005).

### 3.5 Simulation of the model

The Candy Model with parameters inducing highly structured patterns requires adapted Markov Chain Monte Carlo (MCMC) strategies. In this work, we use the Metropolis-Hastings algorithm, which simulates a Markov chain with transition kernel  $P$  that has the Candy point process as its equilibrium or stationary distribution (van Lieshout and Stoica, 2003). The considered transition kernel from the configuration  $\mathbf{x}$  to any  $A \subseteq W \times M$  of  $\mathcal{F}$  (set of subspaces of  $W \times M$ ) writes (e.g., Stoica, 2014)

$$P(\mathbf{x}, A) = \int_A \alpha(\mathbf{x}, \mathbf{y}) Q(\mathbf{x}, d\mathbf{y}) \quad \forall \mathbf{x} \in \Omega, \quad A \in \mathcal{F}, \quad \mathbf{x} \in A,$$

and

$$P(\mathbf{x}, \{\mathbf{x}\}) = \int_{\mathbf{y} \neq \mathbf{x}} [1 - \alpha(\mathbf{x}, \mathbf{y})] Q(\mathbf{x}, d\mathbf{y}),$$

where  $Q$  is the proposal kernel and  $\alpha(\cdot, \cdot)$  is the acceptance probability. This proposal kernel is built from three types of moves (modifications), that are applied to the initial object configuration. These modifications are birth, death and change of an object. For the birth move (with a probability  $p_b$ ), a new segment  $s$  is generated

from the birth proposal density  $b(\mathbf{x}, s)$  and added to the present  $\mathbf{x}$ , then the new configuration  $\mathbf{y} = \mathbf{x} \cup \{s\}$  is accepted with the probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{p_d d(\mathbf{y}, s) p(\mathbf{y})}{p_b b(\mathbf{x}, s) p(\mathbf{x})} \right\}, \quad (12)$$

where  $d(\mathbf{y}, s)$  is the death proposal density. For the death move (with a probability  $p_d$ ), a segment is chosen from the configuration  $\mathbf{x}$  and removed with the probability  $d(\mathbf{x}, s)$ . The new configuration  $\mathbf{y} = \mathbf{x} \setminus \{s\}$  is accepted with the probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{p_b b(\mathbf{x}, s) p(\mathbf{x})}{p_d d(\mathbf{y}, s) p(\mathbf{y})} \right\} \quad (13)$$

Finally, for the change move (with a probability  $p_c$ ), a segment  $s$  is selected from the configuration  $\mathbf{x}$  with the probability  $q(\mathbf{x}, s)$  and undergoes a slight modification of its parameters (either its position, its length or its orientation) with the probability density  $c(\mathbf{x}, s, s')$ . The new configuration  $\mathbf{y} = \mathbf{x} \setminus \{s\} \cup \{s'\}$  is accepted with the probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{q(\mathbf{y}, s') c(\mathbf{y}, s', s) p(\mathbf{y})}{q(\mathbf{x}, s) c(\mathbf{x}, s, s') p(\mathbf{x})} \right\}. \quad (14)$$

This last move is used to improve the behavior of the simulations.

The probabilities  $p_b$ ,  $p_d$  and  $p_c$  are chosen so that  $p_b + p_d + p_c \leq 1$ . The densities  $b(\mathbf{x}, s)$ ,  $d(\mathbf{x}, s)$  and  $c(\mathbf{x}, s, s')$  should be strictly positive to avoid pathological cases. The speed of convergence of the Metropolis-Hastings dynamics is influenced by the choice of the proposals. The choice of uniform laws such as  $b(\mathbf{x}, s) = \frac{1}{\nu(W)}$  and  $d(\mathbf{x}, s) = \frac{1}{n(\mathbf{x})}$  guarantees the necessary convergence properties of the simulated Markov chain (van Lieshout and Stoica, 2003; Stoica, 2014). However, to accelerate the convergence of the chain, it is helpful to use laws that propose more often connected configurations, as achieved by the birth proposal

$$b(\mathbf{x}, s) = (1 - p_a) \cdot \frac{\mathbb{1}\{s \in W \times M\}}{\nu(W)} + p_a \cdot b_a(\mathbf{x}, s), \quad (15)$$

where  $p_a$  is the proposal probability to generate a connected segment and  $b_a(\mathbf{x}, s)$  is a probability density for a segment to be connected and aligned with another segment of the configuration. The latter writes

$$b_a(\mathbf{x}, s) = \frac{1}{n(A(\mathbf{x}))} \sum_{\zeta \in A(\mathbf{x})} \tilde{b}(\zeta, s), \quad (16)$$

where  $A(\mathbf{x})$  is the set of segments of the configuration  $\mathbf{x}$  that do not have all of their extremities connected and are likely to be connected to a new segment, and  $n(A(\mathbf{x}))$  is the number of such segments. Once a segment  $\zeta$  is uniformly selected from the set  $A(\mathbf{x})$ , a new one,  $s = (c_s, a_s, l_s)$  is proposed to be added to the configuration using the density

$$\tilde{b}(\zeta, s) = \frac{\mathbb{1}\{c_s \in \tilde{\mathbf{a}}(\zeta)\} 180^\circ}{\nu(\tilde{\mathbf{a}}(\zeta) \cap W) 2\tau}, \quad (17)$$

where  $\tau$  is the alignment tolerance of Section 3.2, and  $\tilde{\mathbf{a}}(\zeta)$  is the region built from the union of the connection region of the segment  $\zeta$  (centered on its extremities)

which are not containing the extremity of any other segment of the configuration  $\mathbf{x}$ .  $\nu(\tilde{\mathbf{a}}(\zeta) \cap W)$  is the surface area of these connection circles, and  $\mathbb{1}\{\cdot\}$  is the indicator function that selects the segment extremity of the configuration to be connected to the new segment. In (17), the extremity of the new segment is uniformly distributed in  $\tilde{\mathbf{a}}(\zeta)$ , and the orientation  $a_s$  follows a uniform distribution between  $a_\zeta - \tau$  and  $a_\zeta + \tau$  such that  $a_s \in [0^\circ, 180^\circ]$ . This proposal birth density helps the model to propose configurations with connected segments.

## 4 Results

In this section, we employ the Candy Model to demonstrate the interpretation of faults on a  $150 \times 100$  pixel image extracted from the fault likelihood, computed on the three-dimensional Volve data set, as described in Section 2, for which the observation window for the process is taken as  $W = [0, 150] \times [0, 100]$ . In the following, all distances are given relatively to pixel sizes, hence should be multiplied by 9 to give distances in meters (or physical space).

Before applying the Candy Model on the fault likelihood image, we first perform an analysis of the stationarity of the Markov Chain (Section 4.1) built by the Metropolis-Hastings algorithm and its ability to sample the probability distribution (6) while varying the proposal probabilities of choosing birth, death and change moves in the simulation. Secondly, to capture the fault uncertainty and explore the model parameter space associated with the Candy Model, we first show how the geometric parameters (Section 4.2.2) and interaction energy parameters (Section 4.2.3) are chosen and then perform the sampling of various fault network realizations. Level sets (or visit maps) of these fault models enable to communicate and visualize the uncertainty reflected by the output realizations. Finally, Section 4.3 considers a way to rank and compare the model realizations based on another interpretation (Wu et al., 2019) using the empty space function  $F$ .

### 4.1 Simulation dynamics (using only the interaction energy)

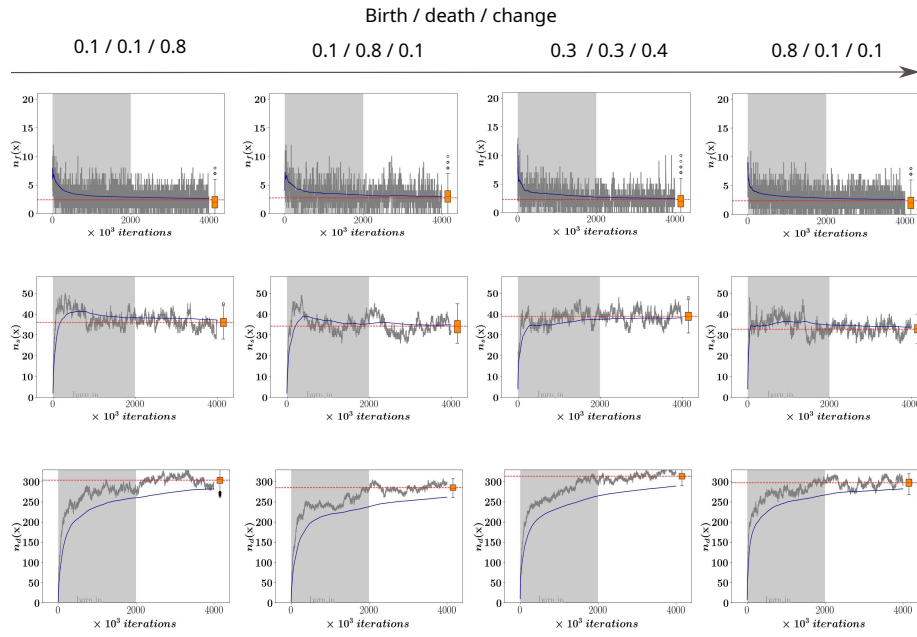
Studying the simulation dynamics of the Candy Model is important to ensure that the simulated Markov chain converges toward the target distribution (Stoica, 2014; van Lieshout and Stoica, 2003). The dynamics of previous versions of the Candy Model has already been studied (van Lieshout and Stoica, 2003; Stoica, 2014; Stoica et al., 2004, 2001). In particular, Reype (2022) performed a sensitivity test on the proposal probabilities ( $p_b$ ,  $p_d$  and  $p_c$ ) of randomly perturbing a configuration of objects. Here, we similarly generate four MCMC chains sampling segments in a window  $W = [0, 100] \times [0, 150]$ , for four sets of proposal probabilities  $p_b$ ,  $p_d$  and  $p_c$ . The segment length and orientations are drawn from prior uniform laws  $\mathcal{U}(14, 19)$  and  $\mathcal{U}(0^\circ, 180^\circ)$ , respectively. For the change moves, we use three types of slight modifications (with a percentage shift  $\Delta_k = 1\%$ ): (1) to slightly move the segment center (with a probability  $p_{move\_pos} = 1/3$ ), (2) to change the orientation of the segment (with a probability  $p_{rotate} = 1/3$ ), and (3) to resize the segment (with a probability  $p_{resize} = 1/3$ ). The proposal probability to perform birth of connected segments is equal to  $p_a = 0.8$  (Eq. (15)). The proposal probabilities are chosen such that, in the first MCMC chain, the birth of segments is

**Table 1** MCMC parameters to simulate the Candy Model.

Parameters	Description	Values
$N_{chains}$	Number of independent simulations.	100
$N_{cycles}$	Number of MCMC cycles. Realizations are kept every cycle.	4000
$N_{moves}$	Number of moves (i.e., transition kernels) in one cycle.	1000
$T$	Initial temperature for accelerating convergence.	5.0
$C$	Cooling schedule for temperature decrease.	0.999
$p_b / p_d / p_c$	Proposal probability for birth/death/change moves in MH (12-13).	0.6 / 0.2 / 0.2
$p_a$	Proposal probability for connected birth for a birth move (15).	0.8
$\Delta k$	Maximum % shift of the segment parameters for a change move.	0.01
$p_{move\_pos}$	Probability to move position.	1/3
$p_{rotate}$	Probability to rotate.	1/3
$p_{resize}$	Probability to resize the segment.	1/3

predominant with  $p_b = 0.8$ ,  $p_d = 0.1$  and  $p_c = 0.1$ . In the second chain, the death of segments is predominant with  $p_b = 0.1$ ,  $p_d = 0.8$  and  $p_c = 0.1$ . In the third chain, the change is slightly predominant with  $p_b = 0.3$ ,  $p_d = 0.3$  and  $p_c = 0.4$ . In the last chain, the strength of the change move is predominant with  $p_b = 0.1$ ,  $p_d = 0.1$  and  $p_c = 0.8$ . For these four chains, the process, in Equation (6), is simulated with a Metropolis-Hastings sampler during  $4 \times 10^6$  iterations corresponding to  $N_{cycles} = 4000$  realizations taken every 1000 iterations. To simulate these chains, we choose  $\log \gamma_f = -3.21$ ,  $\log \gamma_s = -2.56$ ,  $\log \gamma_d = 0.79$  and  $\log \gamma_r = -\infty$ . We consider for the other interaction parameters  $r_c = 1.0$ ,  $\tau = 41.6^\circ$  and  $\delta = 15.0^\circ$ . The burn-in period, considered as part of the chain which is not sampling the target distribution, is empirically fixed to  $2 \times 10^6$  iterations (grey part in Fig. 6).

The sensitivity of the effect of the probability choice  $p_b$ ,  $p_d$  and  $p_c$  is displayed in Fig. 6 in which the sufficient statistics characterizing the model configurations of segments are monitored over MH iterations. The time series of the sufficient statistics (numbers of free, singly and double connected segments) and the corresponding box-plots show that convergence is reached in all cases. The evolution of these sufficient statistics suggests that convergence is reached for all the chains: their evolution tends to approach the same stationary regime in which the post-burn-in estimated means are close, regardless of the configurations of the birth, death and change probabilities. The choice of the probabilities only impacts the convergence speed, as expected theoretically (Stoica, 2014; van Lieshout, 2000; Geyer and Møller, 1994; Møller and Waagepetersen, 2003). We can see that the fastest convergence is obtained when the probability of birth is high. We also notice a good convergence when the probability of choosing a change is slightly higher than the probabilities of choosing a birth or death (Fig. 6, the second last column, with  $p_b = 0.3$ ,  $p_d = 0.3$  and  $p_c = 0.4$ ). Therefore, in the other runs presented in this paper, we choose  $p_b = 0.6$ ,  $p_d = 0.2$  and  $p_c = 0.2$ . In the following sections, we introduce the temperature parameter  $T$ , as in simulated annealing (e.g., Mosegaard and Sambridge, 2002), for which the initial value is set to 5.0, to accelerate the convergence and we consider multiple MCMC chains,  $N_{chains} = 100$ . During the burn-in period,  $T$  slowly decreases (with a rate of  $C = 0.999$ ) toward 1. We thus fix  $2000 \times N_{moves}$  iterations as the burn-in period. Each of the chains is initialized with a set of parameters, for which some are drawn from uniform distributions, see Table 3. All of these MCMC parameters are summarized in Table 1.



**Fig. 6** The simulation time series of sufficient statistics (in gray), cumulative means (in blue), and post burn-in mean number of segments (red dashed line).

## 4.2 Fault network extraction

In this section, we consider, on the one hand, the object parameters defining the segment interactions (Section 3.2) and the data term (10), and, on the other hand, the Candy Model parameters of the Equation (5-6). Simulation results will be analyzed not only through individual realizations, but also through posterior probability level sets to explore the model space associated with the Candy Model.

### 4.2.1 Level sets

To communicate and visualize the uncertainty associated with model parameter space, we consider the level sets of the *a posteriori* probabilities of fault presence. The level sets are often used as mean estimator (Stoica, 2014; Heinrich et al., 2012; Reype, 2022). This can be seen as a scalar field which averages fault network realizations (also called the E-Type by Journel, 1989). In Chapter 7 of Stoica (2014), they are presented to visualize the average shape of the object contained in the data. The objective of this approach is to strengthen the mean while reducing the variance due to fluctuations induced either by the data, the model or by the stochastic method used.

Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_0}$  be realizations sampled from Candy Model of probability density  $p(\mathbf{x}|\theta)$ . Consider a decomposition  $\mathbf{m} = \cup_{k=1}^N m_k$  of the domain  $W$  into a finite number of evenly spaced points  $N$ . For the observed data (fault likelihood image), the grid  $\mathbf{m}$  gives the centers of the pixels and  $c(m_k)$  a grid cell, that is a pixel. In this paper, we consider a mesh size of  $0.25 \times 0.25$ , and a grid origin at



$(0, 0)$ . Let

$$p(m_k) = \mathbb{P}(c(m_k) \in \Phi(\mathbf{x}))$$

where  $\Phi(\mathbf{x}) = \cup_{i=1}^{n(\mathbf{x})} \phi(x_i)$  that is the union of the pixels around a segment that were used to evaluate the data term (Equation (7)). The level set built from the grid points is

$$\ell(m_k) = \{m_k \in \mathbf{m} : p(m_k) > \alpha\}, \quad (18)$$

with  $\alpha \in [0, 1]$ . The estimator of the level set, in case where  $p(m_k)$  is not available, can be approximated by:

$$\hat{p}(m_k) = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbb{1}\{c(m_k) \in \Phi(\mathbf{x}_i)\}, \quad (19)$$

into the level set definition given by Equation (18). The properties of this type of estimator have been studied in Heinrich et al. (2012) and the references therein.

This approach has been used, in this work, to evaluate the quality of fault network extraction and parameter values inducing better network extraction. Here, better network extraction refers to more natural or realistic fault networks. Indeed, according to Stoica (2014), the image obtained by this estimator is a visual average of the shape hidden in the fault probability image. Therefore, this allows us to avoid, at first approximation, any network extraction that may exhibit patterns with “railway-shaped” networks or very complex structures showing too many crossings between segments, resulting in very diffuse regions on the posterior probability map (with wider contours than those of the input fault likelihood image).

#### 4.2.2 Geometric parameters

The networks presented in this paper are obtained by first testing the sensitivity of model parameters, which include geometric parameters (Table 2) and weighting or interaction energy parameters (Table 3). Let us now present the rationale of the geometric parameters. The curvature parameter  $\tau$  defining the orientation between two adjacent segments is encouraged to allow the model to capture curved regions from fault likelihood data. A value of  $30^\circ$  is sufficient for the model to capture these curved regions. The parameter  $\delta$  defining the shadowing interaction in Eq. (4) is taken very wide to eliminate pairs of close and almost parallel segments, which avoids grouping them in the same place. For the connection radius, Lacoste (2004) suggests that it should be less than half the segment length for a good definition of the connection. For instance, choosing too small values may slow down the convergence of the MCMC algorithm (as candidate samples will be rejected more often), whereas greater values may alter the connectivity of simulated patterns. Therefore, we use  $r_c = 1$  as value for the connection radius. Regarding the fault segment length, the literature suggests that it should follow a power law (Bonnet et al., 2001a). In the Candy Model, however, the segment length impacts the shadowing interaction and the curvature of continuous fault segments, and the fault length is primarily controlled by the data. In this paper, we consider an *a priori* non-informative law to see what emerges *a posteriori* through the model. In addition, smaller fault segment length values are suitable for detecting the highly curved areas of the likelihood fault image, whereas greater

**Table 2** Geometric parameters used to simulate the Candy Model.

Parameters	Description	Values
$l$	Segment length.	$\mathcal{U}(11, 14)$
$a$	Segment orientation.	$\mathcal{U}(0^\circ, 180^\circ)$
$\tau$	Maximum orientation for connection interaction.	$30^\circ$
$\delta$	Maximum angle for shadow interacting segments (4).	$85^\circ$
$r_c$	Connection radius.	1.0
$V_{max}$	Energy value of unlikely segments (11).	20
$\sigma$	Fault likelihood threshold (8).	0.3
$\lambda_d$	Scaling factor to balance data and model terms (10).	5.0

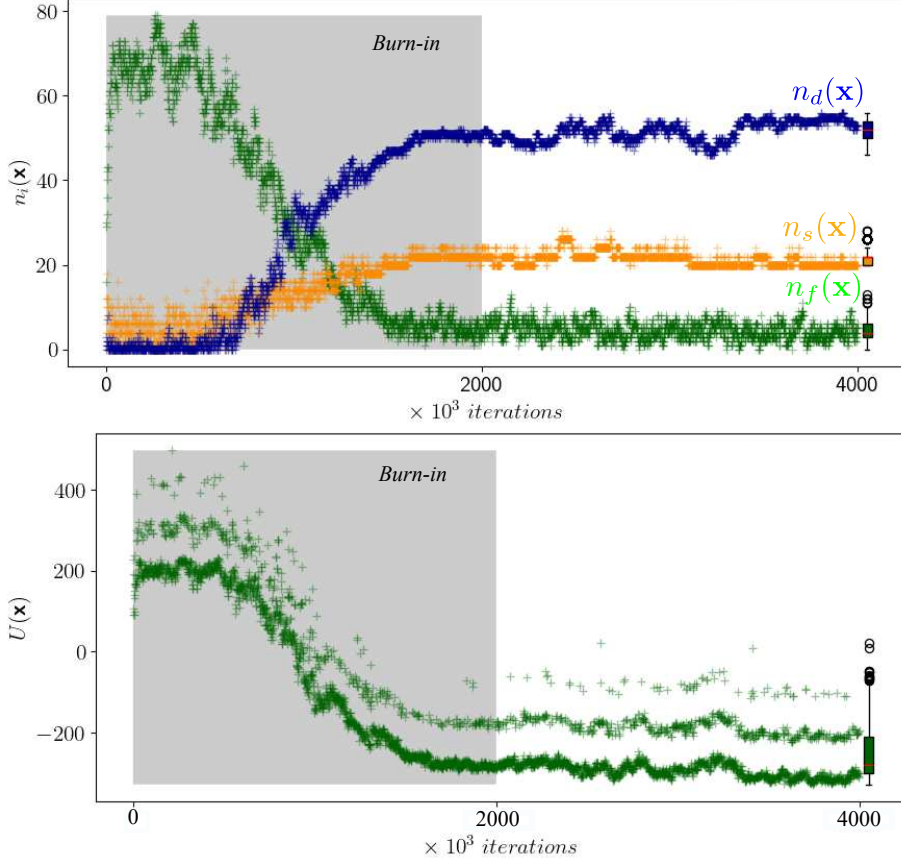
lengths may induce angular configurations and segments reaching areas of low fault likelihood. The segment length is then chosen as following a uniform law  $\mathcal{U}(11, 14)$ . For the segment orientation  $a$ , we consider a non-informative uniform distribution  $\mathcal{U}(0^\circ, 180^\circ)$ , to see whether some patterns emerge as in previous studies (Stoica et al., 2005; Tempel et al., 2014). The parameter  $\lambda_d$  weights the data compared to the *a priori* model (interaction term in Eq. (5)). Preliminary studies showed that for very large values, above  $\lambda_d = 7.5$ , the resulting networks include a lot of intersections (with sub-scale complexity) between the segments along the fault probability corridors. Networks of segments with linear structures then rarely appear even in regions where the probability map suggests so. For very small values (and below  $\lambda_d = 1.0$ ), some regions of the probability map may not be visited by the model, but nevertheless more or less linear and continuous segments are obtained. To filter fault likelihood deemed too low with Eq. (8), we set the fault likelihood threshold, from which the segment data energy is computed, to  $\sigma = 0.3$ . All the above parameters are summarized in table 2.

#### 4.2.3 Interaction energy parameters

The interaction energy parameters are the ones that are hard to calibrate and often call for Bayesian estimation methods (Lacoste, 2004; Stoica et al., 2017). These parameters make it possible to control the morphology of fault networks and can approach the interaction and linkage processes of fault segments well documented in fault studies (Carpenter et al., 2022; Torabi and Berg, 2011; Childs et al., 2009b). We choose to favor the segments connected at the extremities (taking positive values for the log of parameter  $\gamma_d$ ), and to strongly penalize the isolated segments (taking negative values for the log of parameter  $\gamma_f$ ). The log of parameter  $\gamma_s$  is chosen in such a way that it is as negative but greater than  $\log \gamma_f$ . However,  $\log \gamma_s$  should not be very small compared to  $\log \gamma_f$ , in which case, the convergence of the sampling algorithm would be affected, not creating connected segments. To avoid the grouping of fault segments, we choose the ‘hard core’ potential (i.e.,  $\gamma_r = 0$ ), associated with pair of segments reflecting the shadowing interaction. Each of these parameters (meaning  $\log \gamma_f$ ,  $\log \gamma_s$  and  $\log \gamma_d$ ) takes at specific value sampled from uniform laws (Table 3) at the beginning of each MCMC chain.

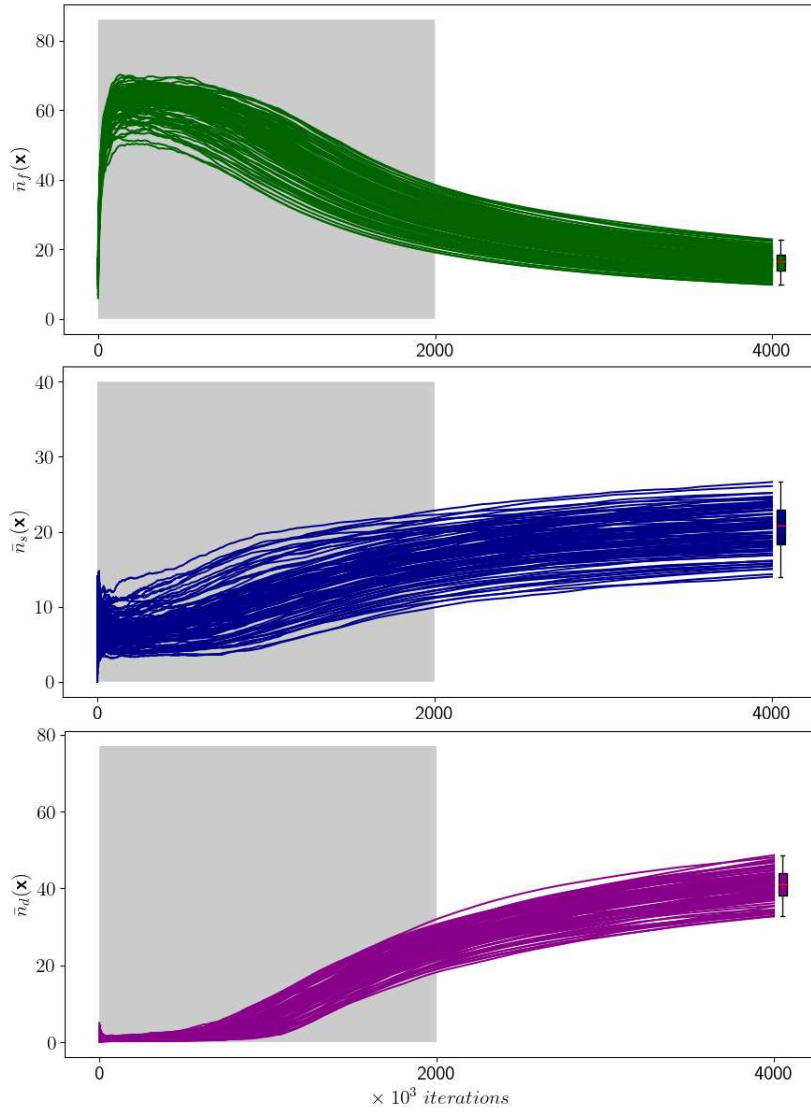
**Table 3** Interaction energy parameters used to simulate the Candy Model (Eq. 5).

Parameters	Description	Values
$\log \gamma_f$	Parameter to control the number of free segments.	$\mathcal{U}(-4.0, -3.0)$
$\log \gamma_s$	Parameter to control the number of singly connected segments.	$\mathcal{U}(-3.0, -2.0)$
$\log \gamma_d$	Parameter to control the number of double connected segments.	$\mathcal{U}(0.1, 1.0)$
$\log \gamma_r$	Parameter to control the number of shadowing segments.	$-\infty$

**Fig. 7** Time series of the number of segments (top), total energy (bottom) and post burn-in boxplots corresponding to the realization in Fig. 9b.

#### 4.2.4 Fault extraction results

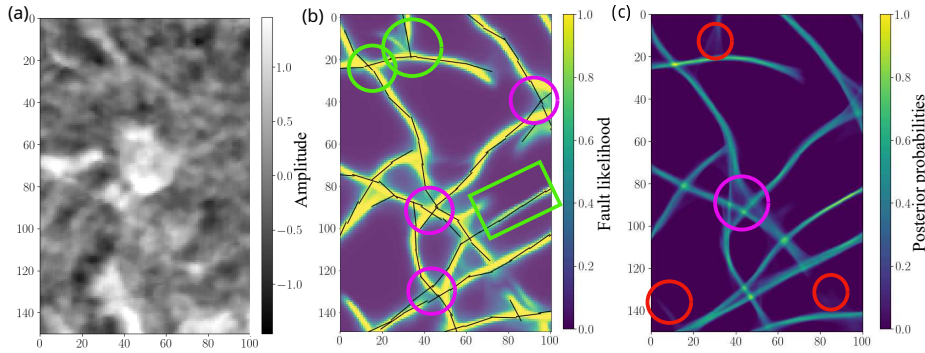
As in Section (4.1), the convergence of a MH simulation chain (Fig. 7) seems to be reached after approximately  $2 \times 10^6$  iterations. When running  $N_{chains} = 100$  simulations, the chains also converge (Fig. 8), but to different values of sufficient statistics owing to the sampling of  $\gamma_d$ ,  $\gamma_s$  and  $\gamma_f$  values. To reflect the uncertainty of these parameters, we consider the ensemble of the results in the remainder of Section 4.2.



**Fig. 8** Cumulated means for free, singly and doubly connected segments (the three panels, respectively) and box-plots computed from the last realizations of  $N_{sim} = 100$  independent simulations (parameters in Tables 2-3).

Fig. 9 shows the fault extraction results obtained with the above set of parameters on a time slice of the Volve seismic image (Fig. 9a).

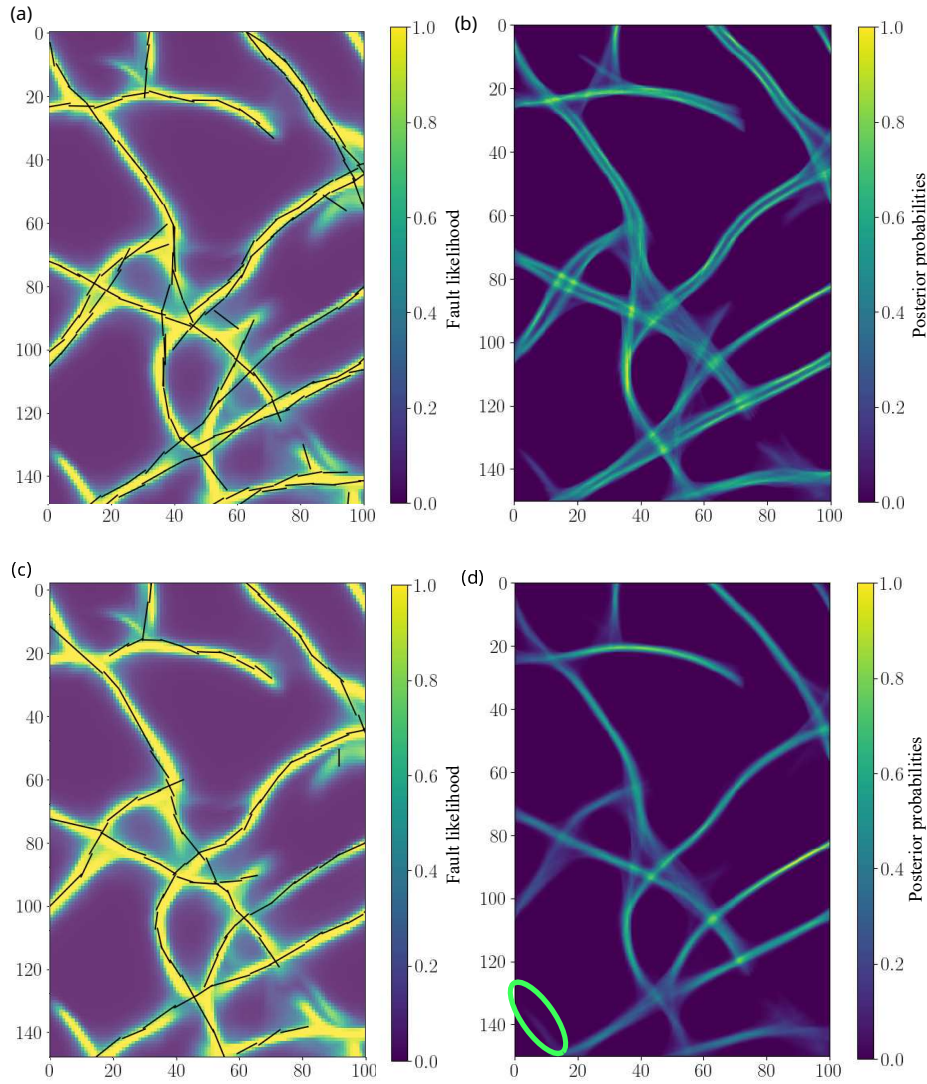
A typical realization (Fig. 9b) manages to extract the main features from the likelihood image. In detail, the main faults are extracted as continuous lines which lie in the likelihood corridors (e.g., green rectangles in Fig. 9b). Branching configurations are also generated by the method (green circles). Finally, the penalization of the shadowing interaction manages to prevent close faults, but it produces some



**Fig. 9** Simulation results (parameters from Tables 1-3). (a)  $150 \times 100$  Pixel seismic time slice; (b) Interpretation result where line segments are in black, and the fault likelihood in the background; (c) The obtained level sets from  $N_{sim} = 100$  simulations. Green rectangles indicate continuous lines of the fault network whereas green and purple circles indicate branching configurations and non-natural fault structures, respectively. Red circles represent areas seldom visited. All of these frames are explained in the text.

“railway-shaped” segment configurations which do not look natural (purple circles in Fig. 9b). This suggests that interactions could be improved in the future versions of the model. The level sets (Fig. 9c) are overall consistent with the input fault likelihood. A focussing is generally visible, which can be explained by the combined effect of the alignment and shadowing interactions, which tends to produce more segments close to the center of the high likelihood corridors than on the edges. We can also note (red circles on Fig. 9c) that some areas of high fault likelihood are seldom visited by the simulated fault segments. These areas are located on the edges of the image and are most often sampled by free or singly connected segments which are penalized by the model parameters (small  $\gamma_f$  and  $\gamma_s$ ).

In the level sets, short segments tend to fill small and highly curved areas with segments. Fig. 10a and c show two examples of realizations while Fig. 10b and d show the level sets obtained when using two other length distributions, one of shorter and narrow range of segment lengths,  $l \sim \mathcal{U}(7, 10)$ , and the other with longer and wider range of segment lengths,  $l \sim \mathcal{U}(5, 50)$ . With shorter segments, the Candy Model generates patterns with “railway-shaped” networks (Fig. 10a) and the image obtained with the level sets (Fig. 10b) indicates a high degree of focussing with these “railway-shape” which is different than in Fig. 9. Using a wider distribution including longer segments, patterns exhibit more linear and continuous structures (Fig. 10c) and the image obtained by the level sets indicates a higher degree of focussing on the fault likelihood corridors (Fig. 10d) as compared to those obtained with the shorter segments. The “railway-shaped” networks no longer appear. For longer segment length range, the model tends to place small segments in highly curved places while in more linear and straight parts, it sometimes places long segments. The difference between the results from the network extractions corresponding to the two prior length distributions comes from the shadowing interaction which highlights a larger shadow zone with the wider range of segment lengths, thus preventing the segments from forming parallel lines.



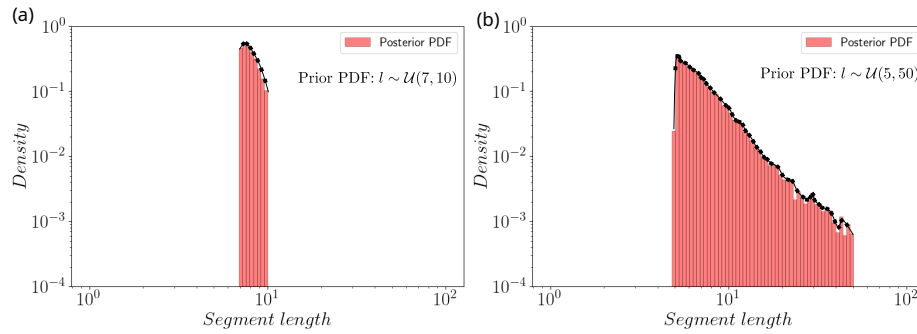
**Fig. 10** Impact of the length distribution. Two segment patterns (a) and (c) and the visit amps corresponding to length distributions  $l \sim \mathcal{U}(7, 10)$  (c) and that of  $l \sim \mathcal{U}(5, 50)$  (d), respectively. Green circles represent areas that are seldom visited, for which explanation is in the text.

#### 4.2.5 Length distribution

We now consider another study on the segment length to assess whether the Candy Model can retrieve realistic length distribution laws even with uniform priors.

Since many of the length distributions analysed in geological fracture (or fault) networks are power distributions (Bonnet et al., 2001b; Davy et al., 2013), we would also like to evaluate the posterior length distributions emerging from the proposed model. In this work, we consider two distributions, one of shorter seg-





**Fig. 11** Impact of the length distribution. Posterior length distribution plot against prior distribution, for prior bounds  $[7, 10]$  (a) and  $[5, 50]$  (b), respectively. The black curve represents the evolution of the distribution with the length size.

ments  $l \sim \mathcal{U}(7, 10)$  and one of longer segments  $l \sim \mathcal{U}(5, 50)$ , representing fault networks at different scales.

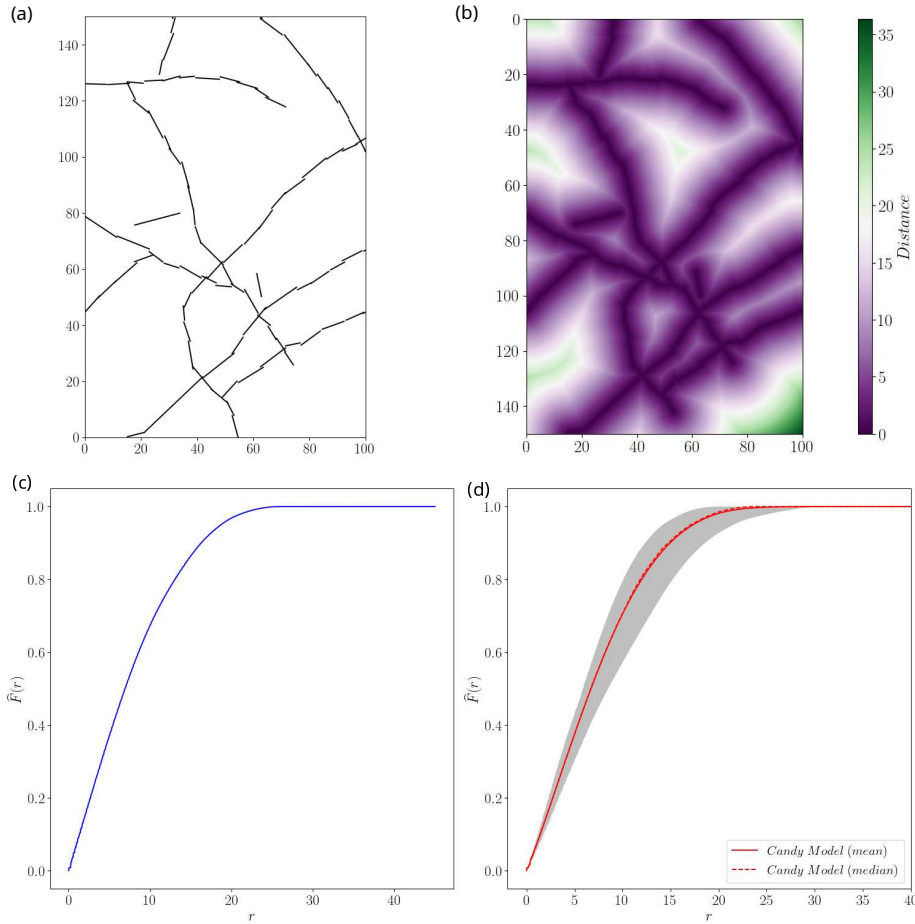
By looking at the posterior length distributions for both prior length distributions, a strong updating is observed (Fig. 11). For the distribution  $l \sim \mathcal{U}(7, 10)$ , the corresponding posterior fault length appears very close to a power law (Fig. 11a). For the distribution  $l \sim \mathcal{U}(5, 50)$ , the corresponding posterior fault length presents small variations (Fig. 11b), however, appears close to a power law. Thus, using non-informative prior distribution, the Candy Model seems to retrieve natural distributions observed for fracture (or fault) networks.

#### 4.3 Analysis and ranking of model realizations

In this section, we consider ways to analyze and rank realizations obtained with the proposed approach. Such a ranking relies on a summary metric, which may be used to compare realizations together, or to compare realizations to some reference solution. The metric used in this work is the empty space function, which originates from stochastic geometry and random set theory (Baddeley et al., 2016). Based on the assumption that the observed pattern is the realization of a stationary process, this method estimates  $F(r) = \mathbb{P}(d(m_k, \mathbf{X}) \leq r)$  where  $m_k$  is a point of the domain, and  $\mathbf{X}$  is the fault process (Kleinschroth et al., 2017; Foxall and Baddeley, 2002; Muru and Tempel, 2023). This function is a valuable way to quantify and test the spatial variability of the interpreted fault segment networks (Kleinschroth et al., 2017; Kruuse et al., 2019). This section uses the empty space function to compare simulated patterns and the skeleton extracted by thinning the fault likelihood (Wu et al., 2019). We test this function on realizations obtained for various parameter sets  $\theta_{i=1, \dots, 5}$  (Table 4), and on the skeleton.

##### 4.3.1 Estimation of the empty space function

The empty space function is defined as the cumulative distribution function of the distances between  $p \in W$  and the closest segment of  $\mathbf{X}$ . This function is approximated from the set of distances sampled on a regular grid (Fig. 12b) using



**Fig. 12** Illustration of the computation of the empty space function. (a) One realization of the Candy Model, (b) the corresponding distance map and (c) the estimated empty space function tested on the Candy Model realization. (d) Range envelopes (gray) encompassing the F functions tested on the last realizations from  $N_{sim} = 100$  simulations, with the parameters of Section 4.2.

the estimator

$$\hat{F}(r) = \frac{\sum_i \mathbb{1}\{d(m_k, W^c) \geq r\} \mathbb{1}\{d(m_k, \mathbf{X}) \leq r\}}{\sum_i \mathbb{1}\{d(m_k, W^c) \geq r\}}, \quad (20)$$

where  $d(m_k, W^c)$  is the shortest distance from a location of a grid point  $m_k \in \mathbf{m}$  to the domain boundary  $W^c$ , so that only image areas which satisfy  $d(s_i, W^c) \geq r$  contribute to the estimate (Baddeley et al., 2016). Here  $\mathbf{m} = \cup_{k=1}^N m_k$  is a decomposition of the domain  $W$  into a finite number of evenly spaced points  $N$  as previously described in Section 4.2. In the following, we consider a mesh size of  $0.2 \times 0.2$ . In a stationary domain  $W$ ,  $\hat{F}(r)$  can be interpreted as the chance that there is a fault segment lying within distance  $r$  of any grid point location.



**Table 4** Parameters used for the ranking of the Candy Model realizations, with the segment length  $l \sim U(11, 14)$ .

$\theta_k$	Interaction energy parameters		
	$\log \gamma_f$	$\log \gamma_s$	$\log \gamma_d$
$\theta_1$	-3.86	-2.56	1.5
$\theta_2$	-3.86	-5.2	1.5
$\theta_3$	-3.1	-2.56	4.0
$\theta_4$	-6.3	-2.56	1.8
$\theta_5$	-4.3	-3.1	1.5

Figure 12d presents an example of  $F$  functions obtained for the envelope of the estimated  $F$  functions for a set of 100 realizations obtained with the same parameters as used for the simulation of faults in Section 4.2. The set of empty space functions can also be summarized by average  $\bar{F}$  and median  $\tilde{F}$  and other summary statistics for any considered distance  $r$ .

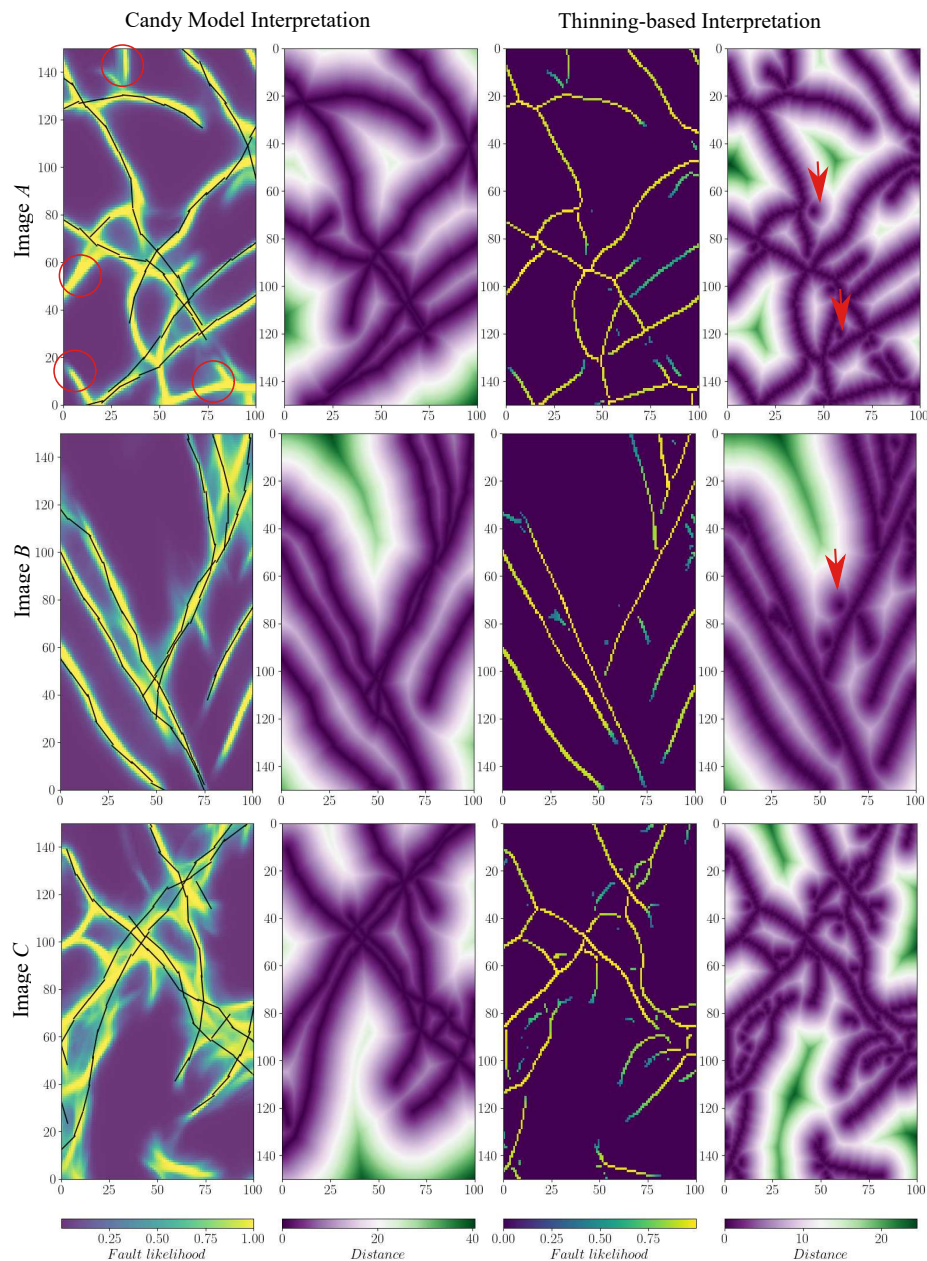
#### 4.3.2 Comparison between the Candy Model realizations and the thinning-based skeleton

The  $F$  function provides a way to compare the spatial variability of several fault segment configurations, which may correspond to realizations of the Candy Model, or to interpretations obtained by other methods. When multiple realizations sampling uncertainty are considered, comparing the envelopes of the  $F$  function to the  $F$  function of some reference solution complements the level sets to characterize results.

In this section, we consider various parameter sets for the Candy Model (Table 4) and use the empty space function envelopes of the corresponding realizations to assess how close they are to the thinning-based fault interpretation of Wu et al. (2019). We might as well choose a manual interpretation made by experts. For representativity, we also consider three sections from the 3D seismic image: two time slices (images  $A$  and  $C$ ) and a crossline (image  $B$ ), see Figure 13. Overall, the segment realizations with parameters  $\theta_1$  display linear connected segments in the thin and elongated high likelihood zones, and intersecting patterns in thicker and diffuse high fault likelihood zones. As in Section 4.2.4, some relatively large likelihood areas are not sampled (red circles in Fig. 13, left column). In contrast, the thinning based interpretation is more regular with visible and continuous fault features.

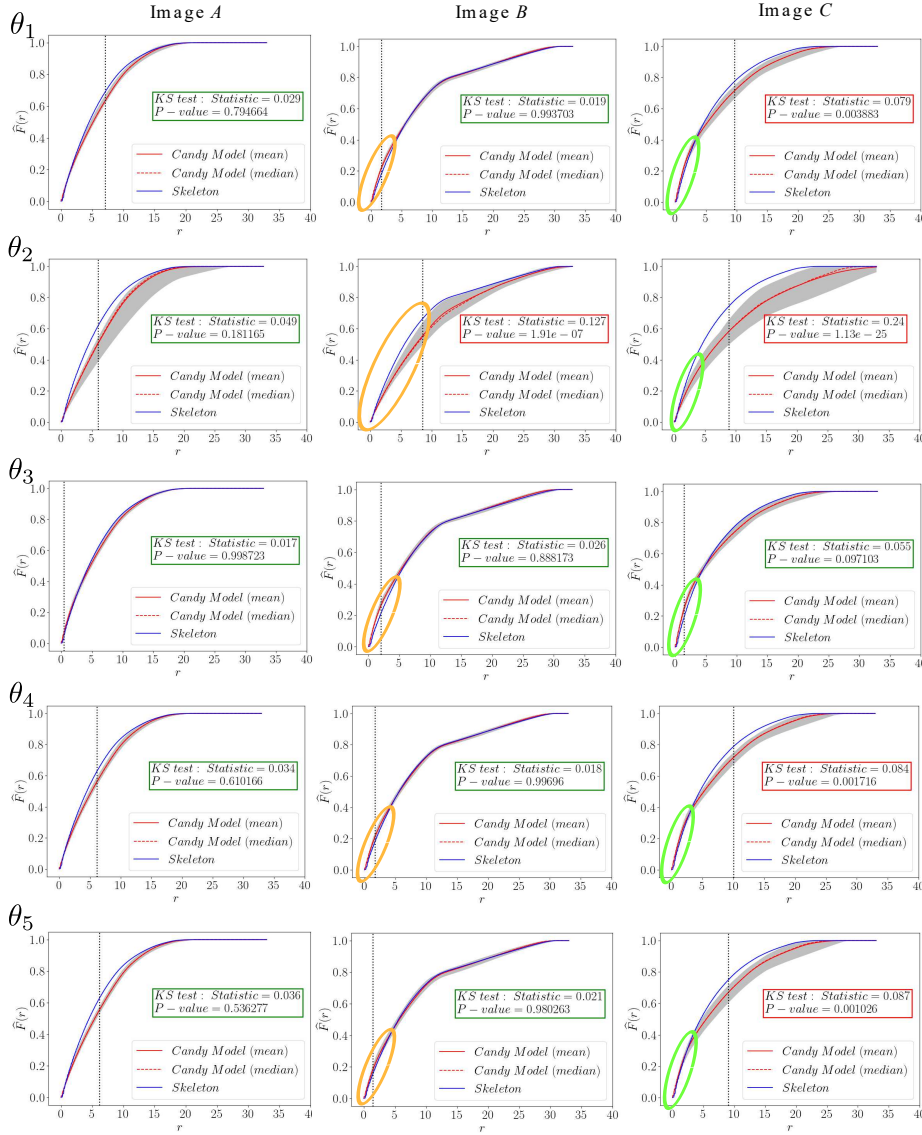
Let us analyze and interpret the results of the empty space function tested on the simulated fault patterns to see the impact of model parameters on  $F$  function envelope estimates and find which model parameter configuration yields realizations with segments spatially organized in a closer way to the skeleton. Fig. 14 shows the empty space functions for the skeletons and the envelopes and means of the empty space functions of the realizations. Range envelopes of the models are diverse.

Figure 14 shows that empty space functions of the skeletons are never wholly included within the envelopes of Candy Model realizations, except for image  $A$  and  $C$  for parameter set  $\theta_1$  and  $\theta_3$  and for images  $B$  for parameter sets  $\theta_1$ ,  $\theta_3$ ,  $\theta_4$ , and  $\theta_5$ . The results of Fig. 14 suggests that the Candy Model generates realizations



**Fig. 13** Results obtained using parameter set  $\theta_1$  on three distinct fault likelihood images. Candy Model realizations (with the lowest energy value) with the corresponding distance maps (left-hand side columns) and the thinning-based skeleton with its distance map (right-hand side columns). Red arrows and circles represent seldom visited areas and isolated small features.

that often have spatial features distinct from those of the skeleton. This difference



**Fig. 14** Empty space functions computed from  $N_{sim} = 100$  simulations with different initial states for parameters  $\theta_{k=1,\dots,5}$ . Circles highlight the differences for smaller values of distances whereas vertical dashed lines indicate where the maximum distance between both distributions, of median empty space function from Candy Model realizations and that of the skeleton, is reached.

is emphasized at shorter distances from encountering a fault segment,  $r < 5$  (i.e., green and orange ellipses in Fig. 14). This is consistent with the visual analysis of distance maps of these two interpretations. Figure 13, which shows that the thinning based process yields several isolated short segments (red arrows in Figure 13) which reduce the empty space as compared to the Candy Model distance fields

(e.g., for image  $A$  with parameters  $\theta_3$ ,  $\theta_4$  and  $\theta_5$  in Fig. 14). In contrast, the Candy Model realizations lack interpreted fault segments at some places (i.e., red circles at second and fifth panels, Fig. 15), thus resulting in an increase of the empty space (corresponding empty space functions for parameter sets  $\theta_3$  and  $\theta_4$  for images  $A$  and  $C$ , Fig. 14). Note that this behavior is reversed for very small distance values (between 2-4 pixels, for parameters  $\theta_1$ ,  $\theta_3$  and  $\theta_4$  for image  $B$  and  $C$ ), which correspond to the presence of some relatively thick fault corridors in the Candy Model realizations (width of 4-8 pixels) in areas where the thinning based process leaves only one linear fault. The results in Fig. 14 are consistent with the Kolmogorov-Smirnov (KS) test (in Appendix A.2) in terms of the values of the statistic. However, for image  $B$  and parameters  $\theta_1$ ,  $\theta_{3-5}$ , the Candy Model realizations seem to fill the space in a way that is quite close to the skeleton. This is consistent with the realizations in the thin and linear fault likelihood corridors in Figure 13.

Overall, the patterns obtained with the parameter set  $\theta_3$  (Fig. 14) appear to produce a relatively narrow envelope and a median function of the empty space closest to that of the skeleton. This means that this parameter set is able to approach the spatial distribution produced by the parsimonious skeleton interpretation.

## 5 Discussion and conclusions

The Candy Model is used to sample from the fault likelihood image as computed by a CNN-based model, FaultNet3D (Wu et al., 2019), to generate various possible fault networks, by modeling them using an explicit formulation, thanks to the marked point process theory. The explicit formulation allows to construct two terms. The first term takes some geological prior knowledge, such as the connectivity at fault segment extremities, formulated as the interaction energy, to favor connected fault networks. The second term takes into account the positions of faults on the fault likelihood image. The framework provided by the Candy Model can be extended by removing or adding terms as needed. In the first application, we have purposely removed terms from the original Candy Model (Stoica et al., 2005), such as those forbidding crossing and  $T$ -shaped interactions, as they are often encountered in geological fault networks (Peacock et al., 2017).

The proposed methodology therein has significant advantages: first it proposes a stochastic modeling approach, which is a point process with interactions, able to model sub-scale and curved faults conditioned to reference data (similar to Barna et al., 2015). The sufficient statistics of the model, representing the number of different types of segment configurations allow to morphologically and quantitatively describe fault networks. When the number of isolated (free) segments is greater, the resulting networks are “fragmented”. When the number of doubly connected segments is greater, this yields “long” networks. The level set estimates of the simulated fault patterns, allows for communicating and visualizing the posterior uncertainty associated with the Candy Model. Finally, the empty space function characterizes the simulated patterns. The analysis carried out in Section 4.3.2 shows that the  $F$  function can be used to rank Candy Model realizations against some interpretation based on spatial variability principles. In the presence of a reference reliable interpretation, this ranking could be used for inferring parameter

values for the Candy Model using Bayesian inference (Bernardo and Smith, 2009; Gelman et al., 1995; Stoica et al., 2017). However, parameter inference remains a challenging task, and calls for considering several criteria: level sets to check for spatial bias and quantify the ability of the Candy Model to filter out possible noise in the likelihood image;  $F$  functions to check the spatial variability features in a statistical sense. Other criteria such as nearest neighbors distance and bivariate  $J$  functions (Baddeley et al., 2016; Foxall and Baddeley, 2002) can also be considered in future studies, but visual checks and assessment by experienced geologists will remain a very important and needed element to judge about the realism of fault networks produced by an algorithm.

The drawback of the proposed methodology can relate to the choice of model parameters, which is one of the key elements of this method since they control the morphology of the generated fault networks. In this paper, we set up our model, data and object parameters on the basis of a sensitivity test analysis and previous works in road detection and deep space imaging and by testing some values guided by the qualitative boxes method (Azencott, 1988). Some parameters may be closely related to the geological problem at stake, features, maps or studied field observations. However, interaction energy parameters interact non linearly, hence they are hard to estimate. Approximated Bayesian Computation (ABC) approaches (Stoica et al., 2017), have been used for effectively addressing this problem. The study carried out in Section 4.2 reveals that the Candy Model can predict realistic length distributions such as power laws (similar to the UFM model, Davy et al., 2010). Future work could consider to tailor interactions to the specificity of the characteristic geometries of segmented faults such as relay zones (e.g., Soliva and Benedicto, 2004; Soliva et al., 2008; Camanni et al., 2019; Childs et al., 2019). To accelerate the convergence of the simulated Markov chain, one could develop adapted proposal laws for such configurations. Available and additional geological information can also be added thanks to the data term. This leaves a large room of perspectives for future work to enrich the proposed approach.

## Acknowledgments

This work was performed in the frame of the RING project at Université de Lorraine (www.ring-team.org). We would like to thank for their support the industrial and academic sponsors of the RING-GOCAD Consortium managed by ASGA. The RING software corresponding to this paper is available as RING-PointProcess, depending on the OpenGeode software (Geode-Solutions, 2022). We also acknowledge Geode-solutions for the OpenGeode open source environment and proprietary tools (geometric, distance objects and grids).

## References

- Ashraf U, Zhang H, Anees A, Nasir Mangi H, Ali M, Ullah Z, Zhang X (2020) Application of Unconventional Seismic Attributes and Unsupervised Machine Learning for the Identification of Fault and Fracture Network. *Appl Sci* 10(11):3864, DOI 10.3390/app10113864

- Aydin O, Caers JK (2017) Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework. *Tectonophysics* 712–713:101–124, DOI 10.1016/j.tecto.2017.04.027
- Azencott R (1988) Image analysis and markov fields. *Proc ICIAM'87* pp 53–61
- Azencott R, Chalmond B, Coldefy F (1995) Markov fusion of a pair of noisy images to detect intensity valleys. *Int J Comput Vision* 16(2):135–145, DOI 10.1007/BF01539552
- Baddeley A, Rubak E, Turner R (2016) *Spatial point patterns: methodology and applications with R*. CRC press
- Barna K, Szirányi T, Borda M, Lavialle O (2015) Marked point processes for enhancing seismic fault patterns. *Journal of Applied Geophysics* 118:115–123, DOI 10.1016/j.jappgeo.2015.04.009
- Bedle H, van Gestel JP (2021) Introduction to this special section: Seismic interpretation. *Lead Edge* 40(7):482–483
- Bense V, Gleeson T, Loveless S, Bour O, Scibek J (2013) Fault zone hydrogeology. *Earth Sci Rev* 127:171–192, DOI 10.1016/j.earscirev.2013.09.008
- Bernardo JM, Smith AF (2009) *Bayesian theory*, vol 405. John Wiley & Sons
- Bond CE (2015) Uncertainty in structural interpretation: Lessons to be learnt. *J Struct Geol* 74:185–200, DOI 10.1016/j.jsg.2015.03.003
- Bonneau F, Caumon G, Renard P (2016) Impact of a stochastic sequential initiation of fractures on the spatial correlations and connectivity of discrete fracture networks. *J Geophys Res: Solid Earth* 121(8):5641–5658
- Bonnet E, Bour O, Odling NE, Davy P, Main I, Cowie P, Berkowitz B (2001a) Scaling of fracture systems in geological media. *Rev Geophys* 39(3):347–383
- Bonnet E, Bour O, Odling NE, Davy P, Main I, Cowie P, Berkowitz B (2001b) Scaling of fracture systems in geological media. *Rev Geophys* 39(3):347–383, DOI 10.1029/1999RG000074
- Botter C, Cardozo N, Lecomte I, Rotevatn A, Paton G (2017) The impact of faults and fluid flow on seismic images of a relay ramp over production time. *Pet Geosci* 23(1):17–28, DOI 10.1144/petgeo2016-027
- Bresenham JE (1965) Algorithm for computer control of a digital plotter. *IBM Syst J* 4(1):25–30
- Camanni G, Roche V, Childs C, Manzocchi T, Walsh J, Conneally J, Saqab MM, Delogkos E (2019) The three-dimensional geometry of relay zones within segmented normal faults. *J Struct Geol* 129:103895, DOI 10.1016/j.jsg.2019.103895
- Carpenter M, Williams JN, Åke Fagereng, Wedmore LN, Biggs J, Mphepo F, Mdala H, Dulanya Z, Manda B (2022) Comparing intrarift and border fault structure in the malawi rift: Implications for normal fault growth. *J Struct Geol* 165:104761, DOI <https://doi.org/10.1016/j.jsg.2022.104761>
- Cherpeau N, Caumon G (2015) Stochastic structural modelling in sparse data situations. *Pet Geosci* 21(4):233–247, DOI 10.1144/petgeo2013-030
- Cherpeau N, Caumon G, Lévy B (2010) Stochastic simulations of fault networks in 3D structural modeling. *CR Geosci* 342(9):687–694, DOI 10.1016/j.crte.2010.04.008
- Childs C, Manzocchi T, Walsh JJ, Bonson CG, Nicol A, Schöpfer MP (2009a) A geometric model of fault zone and fault rock thickness variations. *J Struct Geol* 31(2):117–127
- Childs C, Manzocchi T, Walsh JJ, Bonson CG, Nicol A, Schöpfer MP (2009b) A geometric model of fault zone and fault rock thickness variations. *J Struct Geol*

- 31(2):117–127, DOI 10.1016/j.jsg.2008.08.009
- Childs C, Worthington RP, Walsh JJ, Roche V (2019) Conjugate relay zones: geometry of displacement transfer between opposed-dipping normal faults. *J Struct Geol* 118:377–390, DOI 10.1016/j.jsg.2018.11.007
- Chiu SN, Stoyan D, Kendall WS, Mecke J (2013) *Stochastic geometry and its applications*, 3rd edn. Wiley series in probability and statistics, John Wiley & Sons Inc, Chichester, West Sussex, United Kingdom
- Chopra S, Castagna J, Portniaguine O (2006) Seismic resolution and thin-bed reflectivity inversion. *CSEG recorder* 31(1):19–25
- Das V, Pollack A, Wollner U, Mukerji T (2019) Convolutional neural network for seismic impedance inversion. *Geophysics* 84(6):R869–R880
- Davy P, Le Goc R, Darcel C, Bour O, De Dreuzy JR, Munier R (2010) A likely universal model of fracture scaling and its consequence for crustal hydromechanics. *Journal of Geophysical Research: Solid Earth* 115(B10):2009JB007043, DOI 10.1029/2009JB007043
- Davy P, Le Goc R, Darcel C (2013) A model of fracture nucleation, growth and arrest, and consequences for fracture density and scaling. *Journal of Geophysical Research: Solid Earth* 118(4):1393–1407, DOI 10.1002/jgrb.50120
- Descobes X (ed) (2012) *Stochastic geometry for image analysis (Digital Signal and Image Processing Series)*. John Wiley & Sons
- Dimmen V, Rotevatn A, Lecomte I (2023) Imaging of small-scale faults in seismic reflection data: Insights from seismic modelling of faults in outcrop. *Mar Pet Geol* 147:105980, DOI 10.1016/j.marpetgeo.2022.105980
- Faleide TS, Braathen A, Lecomte I, Mulrooney MJ, Midtkandal I, Bugge AJ, Planke S (2021a) Impacts of seismic resolution on fault interpretation: Insights from seismic modelling. *Tectonophysics* 816:229008, DOI <https://doi.org/10.1016/j.tecto.2021.229008>
- Faleide TS, Braathen A, Lecomte I, Mulrooney MJ, Midtkandal I, Bugge AJ, Planke S (2021b) Impacts of seismic resolution on fault interpretation: Insights from seismic modelling. *Tectonophysics* 816:229008, DOI 10.1016/j.tecto.2021.229008
- Foxall R, Baddeley A (2002) Nonparametric measures of association between a spatial point process and a random set, with geological applications. *J R Stat Soc C* 51(2):165–182
- Froner B, Purves S, Lowell J, Henderson J (2013) Perception of visual information: the role of colour in seismic interpretation. *First Break* 31(4), DOI 10.3997/1365-2397.2013010
- Gabriel AA, Ampuero JP, Dalguer LA, Mai PM (2013) Source properties of dynamic rupture pulses with off-fault plasticity. *J Geophys Res: Solid Earth* 118(8):4117–4126, DOI 10.1002/jgrb.50213
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian data analysis*. Chapman and Hall/CRC
- Geode-Solutions (2022) OpenGeode framework. DOI 10.5281/ZENODO.3610370
- Geyer CJ, Møller J (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand J Stat* pp 359–373
- Godefroy G, Caumon G, Laurent G, Bonneau F (2021) Multi-scenario Interpretations From Sparse Fault Evidence Using Graph Theory and Geological Rules. *J Geophys Res: Solid Earth* 126(2), DOI 10.1029/2020JB020022



- Goodwin H, Aker E, Røe P (2022) Stochastic Modeling of Subseismic Faults Conditioned on Displacement and Orientation Maps. *Math Geosci* 54(1):207–224, DOI 10.1007/s11004-021-09965-7
- Hale D (2013) Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. *Geophysics* 78(2):O33–O43
- Heinrich P, Stoica RS, Tran VC (2012) Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics* 2:47–61, DOI 10.1016/j.spasta.2012.10.001
- Holden L, Mostad P, Nielsen BF, Gjerde J, Townsend C, Ottesen S (2003) Stochastic structural modeling. *Math Geol* 35(8):899–914, DOI <https://doi.org/10.1023/B:MATG.0000011584.51162.69>
- Iacopini D, Butler RW (2011) Imaging deformation in submarine thrust belts using seismic attributes. *Earth Planet Sci Lett* 302(3-4):414–422, DOI 10.1016/j.epsl.2010.12.041
- Imran QS, Siddiqui NA, Latiff AHA, Bashir Y, Khan M, Qureshi K, Al-Masgari AAS, Ahmed N, Jamil M (2021) Automated Fault Detection and Extraction under Gas Chimneys Using Hybrid Discontinuity Attributes. *Appl Sci* 11(16):7218, DOI 10.3390/app11167218
- Islam MR, Shinjo R (2009) Mining-induced fault reactivation associated with the main conveyor belt roadway and safety of the Barapukuria Coal Mine in Bangladesh: Constraints from BEM simulations. *Int J Coal Geol* 79(4):115–130, DOI 10.1016/j.coal.2009.06.007
- Jacquemin P, Mallet JL (2005) Automatic faults extraction using double hough transform. In: SEG International Exposition and Annual Meeting, SEG, pp SEG–2005
- Journel A (1989) Fundamentals of geostatistics in five lessons, short course geol. AGU, Washington, DC
- Julio C, Caumon G, Ford M (2015) Sampling the uncertainty associated with segmented normal fault interpretation using a stochastic downscaling method. *Tectonophysics* 639:56–67, DOI 10.1016/j.tecto.2014.11.013
- Kleinschroth F, Healey JR, Gourlet-Fleury S, Mortier F, Stoica RS (2017) Effects of logging on roadless space in intact forest landscapes of the Congo Basin. *Conserv Biol* 31(2):469–480
- Kruuse M, Tempel E, Kipper R, Stoica RS (2019) Photometric redshift galaxies as tracers of the filamentary network. *A & A* 625:A130, DOI 10.1051/0004-6361/201935096
- Lacoste C (2004) Extraction de réseaux linéiques à partir d'images satellitaires et aériennes par processus ponctuels marqués. PhD thesis, Université Nice Sophia Antipolis
- Lacoste C, Descombes X, Zerubia J, Baghdadi N (2004) Bayesian geometric model for line network extraction from satellite images. In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, vol 3, pp iii–565
- Lafarge F, Descombes X (2012) Structure extraction. In: Descombes X (ed) *Stochastic geometry for image analysis (Digital Signal and Image Processing Series)*, John Wiley & Sons, chap 8, pp 220–255
- van Lieshout M (2000) Markov point processes and their applications. World Scientific
- van Lieshout MNM, Stoica RS (2003) The Candy model: properties and inference. *Stat Neerl* 57(2):177–206, DOI 10.1111/1467-9574.00227



- Lopes RH, Hobson PR, Reid ID (2008) Computationally efficient algorithms for the two-dimensional kolmogorov–smirnov test. *J Phys Conf Ser* 119(4):042019
- Mabon J, Ortner M, Zerubia J (2023) Apprentissage contrastif de modèles de processus ponctuels pour la détection d’objets. In: *GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images*, Grenoble, France
- Manzi MS, Cooper GR, Malehmir A, Durrheim RJ (2020) Improved structural interpretation of legacy 3D seismic data from Karee platinum mine (South Africa) through the application of novel seismic attributes. *Geophys Prospect* 68(1):145–163, DOI 10.1111/1365-2478.12900
- Manzocchi T, Walsh JJ, Nell P, Yielding G (1999) Fault transmissibility multipliers for flow simulation models. *Pet Geosci* 5(1):53–63, DOI 10.1144/petgeo.5.1.53
- Marfurt KJ, Kirlin RL, Farmer SL, Bahorich MS (1998) 3-D seismic attributes using a semblance-based coherency algorithm. *GEOPHYSICS* 63(4):1150–1165, DOI 10.1190/1.1444415
- Moller J, Waagepetersen RP (2003) *Statistical inference and simulation for spatial point processes*. CRC press
- Mosegaard K, Sambridge M (2002) Monte Carlo analysis of inverse problems. *Inverse Probl* 18(3):R29–R54, DOI 10.1088/0266-5611/18/3/201
- Muru MM, Tempel E (2023) Using photometric redshift data to improve the detection of galactic filaments with the Bisous model. *ArXiv:2301.02710 [astro-ph]*
- Peacock DCP, Nixon C, Rotevatn A, Sanderson D, Zuluaga L (2017) Interacting faults. *J Struct Geol* 97:1–22
- Ravasi M, Vasconcelos I, Curtis A, Kritski A (2015) Vector-acoustic reverse time migration of Volve ocean-bottom cable data set without up/down decomposed wavefields. *GEOPHYSICS* 80(4):S137–S150, DOI 10.1190/geo2014-0554.1
- Reype C (2022) *Probabilistic modelling and bayesian inference for the analysis of geological fluid mixing systems: pattern detection and parameter estimation*. Phd thesis, Université de Lorraine
- Roche V, Camanni G, Childs C, Manzocchi T, Walsh J, Conneally J, Saqab MM, Delogkos E (2021) Variability in the three-dimensional geometry of segmented normal fault surfaces. *Earth Sci Rev* 216:103523, DOI 10.1016/j.earscirev.2021.103523
- Rosenbaum MS, Culshaw MG (2003) Communicating the Risks Arising from Geohazards. *J R Stat Soc A* 166(2):261–270, DOI 10.1111/1467-985X.00275
- Soleng H, Rivenæs J, Gjerde J, Hollund K, Holden L (2004) Structural Uncertainty Modelling and the Representation of Faults as Staircases. In: *EC-MOR IX - 9th European Conference on the Mathematics of Oil Recovery*, European Association of Geoscientists & Engineers, Cannes, France,, DOI 10.3997/2214-4609-pdb.9.A002
- Soliva R, Benedicto A (2004) A linkage criterion for segmented normal faults. *J Struct Geol* 26(12):2251–2267
- Soliva R, Benedicto A, Schultz R, Maerten L, Micarelli L (2008) Displacement and interaction of normal fault segments branched at depth: Implications for fault growth and potential earthquake rupture size. *J Struct Geol* 30(10):1288–1299
- STATOIL (1993) DISCOVERY EVALUATION REPORT. Tech. rep., Sleipner Petek, URL [https://datavillagesa.blob.core.windows.net/volve/Reports/Discovery\\_report.pdf](https://datavillagesa.blob.core.windows.net/volve/Reports/Discovery_report.pdf)
- Stoica RS (2001) *Processus Ponctuels pour l’Extraction de R seaux Lin ques dans les Images Satellitaires et A riennes*. PhD thesis, Université de Lorraine

- Stoica RS (2014) Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées. Habilitation à diriger des recherches en mathématiques, l'Université Lille 1, 232 pages
- Stoica RS, Descombes X, Zérubia J (2001) Road extraction in remote sensed images using stochastic geometry framework. *AIP Conf Proc* 568(1):531–542
- Stoica RS, Descombes X, Zerubia J (2004) A Gibbs Point Process for Road Extraction from Remotely Sensed Images. *Int J Comput Vision* 57(2):121–136, DOI 10.1023/B:VISI.0000013086.45688.5d
- Stoica RS, Martínez VJ, Mateu J, Saar E (2005) Detection of cosmic filaments using the Candy model. *A & A* 434(2):423–432, DOI 10.1051/0004-6361:20042409
- Stoica RS, Gay E, Kretschmar A (2007) Cluster Pattern Detection in Spatial Data Based on Monte Carlo Inference. *Biom J* 49(4):505–519, DOI 10.1002/bimj.200610326
- Stoica RS, Philippe A, Gregori P, Mateu J (2017) ABC Shadow algorithm: a tool for statistical analysis of spatial patterns. *Stat Comput* 27(5):1225–1238, DOI 10.1007/s11222-016-9682-x
- Stoica RS, Deaconu M, Philippe A, Hurtado-Gil L (2021) Shadow Simulated Annealing: A new algorithm for approximate Bayesian inference of Gibbs point processes. *Spatial Stat* 43:100505, DOI 10.1016/j.spasta.2021.100505
- Szydluk TJ, Way S, Smith P, Aamodt L, Friedrich C (2006) 3d pp/ps prestack depth migration on the volve field. In: 68th EAGE Conference and Exhibition incorporating SPE EUROPEC 2006, EAGE Publications BV, pp cp-2
- Tempel E, Stoica RS, Martínez VJ, Liivamägi LJ, Castellan G, Saar E (2014) Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. *Mon Not R Astron Soc* 438(4):3465–3482, DOI 10.1093/mnras/stt2454
- Thore P, Shtuka A, Lecour M, Ait-Ettajer T, Cognot R (2002) Structural uncertainties: Determination, management, and applications. *Geophysics* 67(3):840–852
- Torabi A, Berg SS (2011) Scaling of fault attributes: A review. *Mar Pet Geol* 28(8):1444–1460, DOI 10.1016/j.marpetgeo.2011.04.003
- Tunkiel AT, Wiktorski T, Sui D (2020) Drilling Dataset Exploration, Processing and Interpretation Using Volve Field Data. In: Volume 11: Petroleum Technology, American Society of Mechanical Engineers, Virtual, Online, p V011T11A076, DOI 10.1115/OMAE2020-18151
- Wu X, Shi Y, Fomel S, Liang L, Zhang Q, Yusifov AZ (2019) FaultNet3D: Predicting Fault Probabilities, Strikes, and Dips With a Single Convolutional Neural Network. *IEEE Trans Geosci Remote Sens* 57(11):9138–9155, DOI 10.1109/TGRS.2019.2925003
- Wu X, Geng Z, Shi Y, Pham N, Fomel S, Caumon G (2020) Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *GEOPHYSICS* 85(4):WA27–WA39, DOI 10.1190/geo2019-0375.

## Appendix A Supplementary materials

### Appendix A.1 algorithm for simulating the Candy Model

In this paper, the Metropolis-Hastings (Algorithm 1) allows to sample the distribution of the Candy Model.

---

#### Algorithm 1 Candy Model simulation with Metropolis-Hastings

---

**Require:**

$N_{cycles}$  : Number of MCMC cycles. Realizations are kept every cycle.  
 $N_{moves}$  : Number of moves in one cycle (i.e., transition kernels).  
 $p_b/p_d/p_c$  : Proposal probability for birth/death/change moves in MH.  
 $p_a$  : Proposal probability for connected birth for a birth move.  
 $p_{move\_pos}$  : Probability to move position.  
 $p_{rotate}$  : Probability to rotate.  
 $p_{resize}$  : Probability to resize the segment.  
 $\Delta k$  : Maximum % shift of the segment parameters for a change move.  
 $\mathbf{x}$  : Configuration of the segments (initialized empty).  
 $p(\cdot|\theta)$  : Probability distribution of the considered model.

```

while  $N_{cycles} \neq 0$  do
  while  $N_{moves} \neq 0$  do
     $u \sim \mathcal{U}(0, 1)$ 
    if  $u < p_b$  then ▷ Propose a birth of a segment
      Sample segment  $s$  according to  $b(\mathbf{x}, s)$  from Equation (15)
       $w \sim \mathcal{U}(0, 1)$ 
      if  $w < \min\{1, \alpha(\mathbf{x}, \mathbf{x} \cup \{s\})\}$  then ▷ with  $\alpha$  from Equation (12)
         $\mathbf{x} \leftarrow \mathbf{x} \cup \{s\}$ 
      end if
    else if  $u < p_b + p_d$  then ▷ Propose a death of a segment
      Select segment  $s$  according to  $d(\mathbf{x}, s)$  in Section (refsec:3.5)
       $w \sim \mathcal{U}(0, 1)$ 
      if  $w < \min\{1, \alpha(\mathbf{x}, \mathbf{x} \setminus \{s\})\}$  then ▷ with  $\alpha$  from Equation (13)
         $\mathbf{x} \leftarrow \mathbf{x} \setminus \{s\}$ 
      end if
    else ▷ Propose a slight ( $\Delta k$ ) change of a segment
      Select segment  $s$  according to  $d(\mathbf{x}, s)$  (Section 3.5)
      Either move, rotate or stretch  $s$  according to  $c(\mathbf{x}, s, s')$  (Section 3.5)
       $w \sim \mathcal{U}(0, 1)$ 
      if  $w < \min\{1, \alpha(\mathbf{x}, \mathbf{x} \setminus \{s\} \cup \{s'\})\}$  then ▷ with  $\alpha$  from Equation (14)
         $\mathbf{x} \leftarrow \mathbf{x} \setminus \{s\} \cup \{s'\}$ 
      end if
    end if
     $N_{moves} \leftarrow N_{moves} - 1$ 
  end while
  Update  $\mathbf{x}$  in the chain.
   $N_{cycles} \leftarrow N_{cycles} - 1$ 
end while

```

---

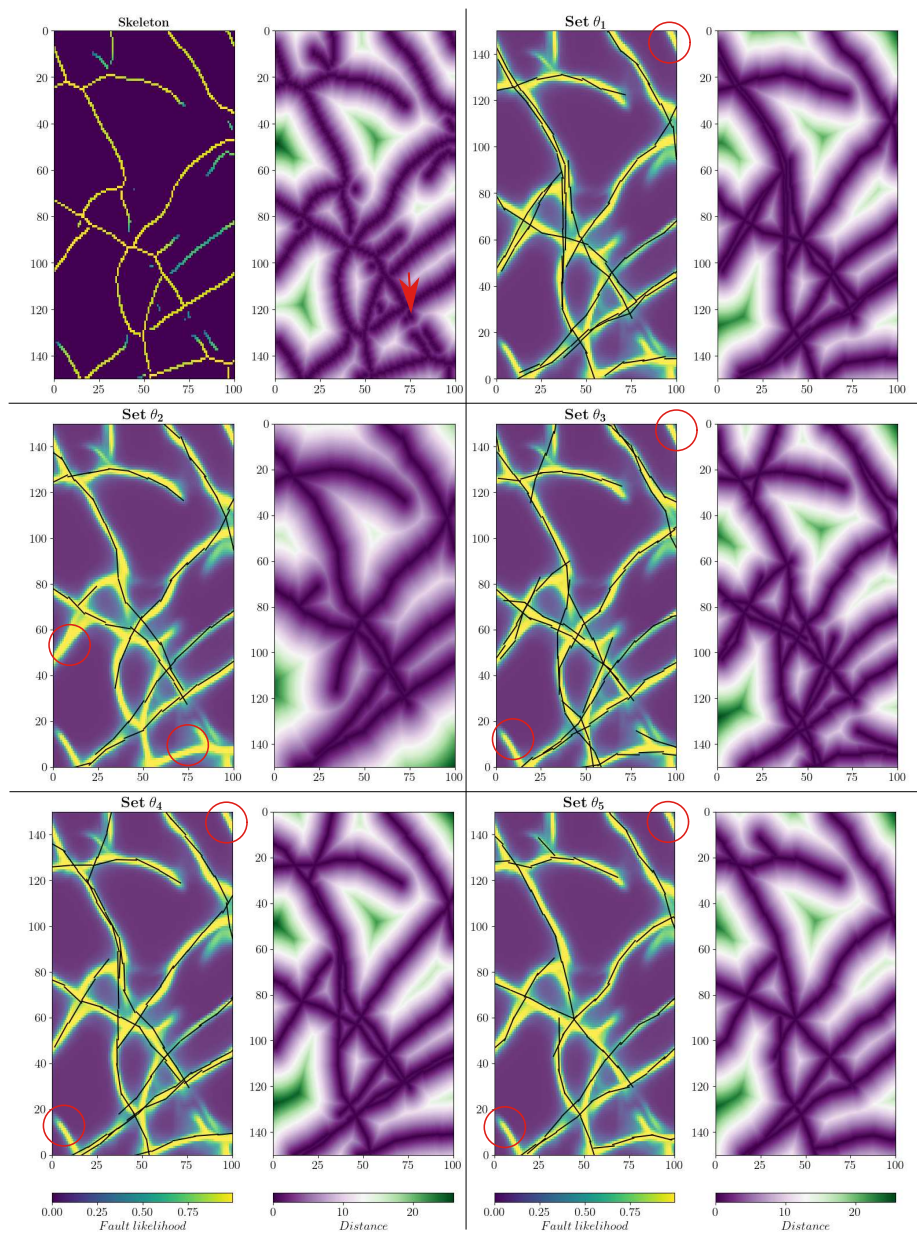
### Appendix A.2 Comparison between Candy Model realizations and the thinning-based skeleton through the Kolmogorov-Smirnov (KS) test

In complement to the analysis carried out in Section 4.3.2, the Kolmogorov-Smirnov (KS) test is used for comparing the Candy realizations and the thinning-based interpretation. This is a

non-parametric statistic, designed for comparing two empirical distributions, by defining the largest absolute difference between the two cumulative distribution functions as a measure of disagreement (Lopes et al., 2008). The null hypothesis states that the samples  $\tilde{F}_{\theta_k}(j)$  and  $\hat{F}_{skeleton}(j)$ , for  $j = 1, \dots, n$ , are drawn from the same distribution. We choose a confidence level of 95%, that is, we will reject the null hypothesis in favor of the alternative if the  $p$ -value is less than 0.05.

The analysis for KS test is carried out using the `scipy.stats.ks_2samp` function from the `scipy` library. This test performs the two-sample KS test of goodness of fit, which compares the underlying distributions of two independent samples from the continuous empty space functions of the Candy realizations and the skeleton. The test is carried out on three images (Fig. 13), each one using five parameters (Table 4), in which  $N_{sim} = 100$  simulations are run. The comparison between both interpretations is done for each parameter set, through the median  $\tilde{F}_{\theta_k}$  curves obtained for each model parameter set  $\theta_k$  with the empty space  $\hat{F}_{skeleton}$  of the skeleton.

Figure 14 shows the results of the Kolmogorov-Smirnov statistical test on boxes in either green or red. For the *A* image, the  $p$ -values of all parameter sets are greater than 0.05, which means that the null hypothesis is not rejected. This is consistent with the results of Fig. 14 even though envelope test seems to be more discriminating. Actually, envelopes show how the realizations of the Candy Model from such an image fill the space in a way quite close to the skeleton process. For image *B*, only the  $p$ -value corresponding to the parameter set  $\theta_2$  is less than 0.05, indicating that Candy Model realization that fill the space, in a way quite differently than the skeleton for this image (which is also consistent with the two other images). For image *C*, only the  $p$ -value of the parameter set  $\theta_3$  is greater than 0.05, which indicates the parameter set inducing realizations filling the space in a closer way as the skeleton.



**Fig. 15** Interpretation on image A. Interpreted fault networks (first column) with the lowest energy value and their distance maps (second column). Parameter set  $\theta_{i=1,\dots,5}$  are those Table 4. The red arrows and circles indicate isolated features in the skeleton and seldom visited areas in the Candy Model realizations, respectively.