Hilde Grude Borgos

Stochastic Modeling and Statistical Inference of Geological Fault Populations and Patterns

Dr. Ing. Thesis

Department of Mathematical Sciences
Norwegian University of Sciences and Technology
2000

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree "Doktor Ingeniør" (Dr.Ing.) at the Norwegian University of Science and Technology (NTNU). The work is financed by a scholarship from VISTA, a research corporation between The Norwegian Academy of Science and Letters and Den norske stats oljeselskap a.s. (Statoil).

I first of all thank my supervisor Henning Omre at the Department of Mathematical Sciences for all guidance and support. I also thank my cosupervisor Chris Townsend at Statoil R&D for helpful advises on the geological aspects of my research and for providing data. I thank my coauthors Patience Cowie and Nancye Dawers for a fruitful collaboration. I am especially grateful to Patience Cowie for inviting me to stay six months at the Department of Geology and Geophysics, the University of Edinburgh, UK, during the academic year 1998/99, and would like to thank all my colleagues during my stay in Scotland for making this winter memorable. I thank Noelle Odling at Nansen Environmental and Remote Sensing Center for permission to use data from Hornelen, western Norway. A special thanks to the Department of Mathematical Sciences at NTNU, in particular my colleagues in the statistics group and also the administrative staff, for providing a stimulating working environment. The warmest thanks goes to Jan Terje for all love and support.

Trondheim, May 2000

Hilde Grude Borgos

Introduction

The content of this thesis is listed below. A background for the thesis is presented, giving an introduction to the concepts of faults, fault analysis and modeling to readers not familiar with this subject. A summary of the thesis is also provided.

Thesis Content

The thesis consists of the following units:

Part I: **Model choice for fault size distributions**
Report.

Part II: **Sampling algorithm for estimating Bayes factor**
Report.

Part III: **Practicalities of extrapolating 1D fault and fracture size-frequency distributions to higher dimensional samples**
With Patience A. Cowie and Nancye H. Dawers.
Paper to appear in *Journal of Geophysical Research - Solid Earth.*

Part IV: **Stochastic model for fault geometry conditioned to seismic data and well observations**
Report.

Part V: **Partitioning of a line segment**
Report.

Appendix A: **Model choice for fault distribution**
With Henning Omre.
Abstract; Conference: Modelling Permeable Rocks, Cambridge, UK, March 1998.

Appendix B: **Stochastic simulation of fault patterns conditioned on seismic data and well data**
Extended abstract; Conference: EAGE/SPE international symposium on petroleum geostatistics, Toulouse, France, April 1999.

Appendix C: **Uncertainty in fault geometries**
With Henning Omre.
Paper to appear in Proceedings of the 6th International Geostatistics Congress, Cape Town, South Africa, April 2000.

The units can be read independently of each other. It is however recommended to read Part I, II and III, which constitute the main part of the thesis, in this respective order. Part IV is an independent part not related to the previous units. Part V is a short note. Part III is written for a geophysical audience. The remaining parts are written for a statistical audience, but can also be read by people with an interest in quantitative geology. The appendices are abstracts and papers based on Part I and IV.

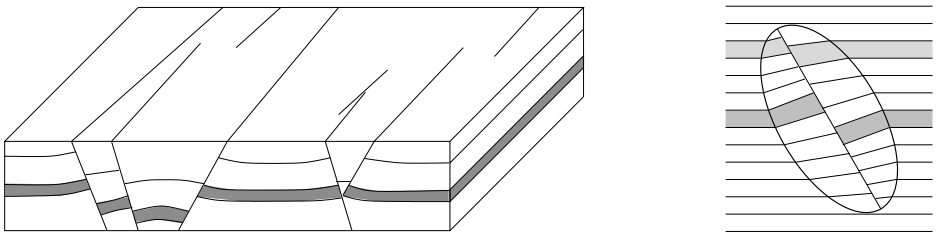
Thesis Background

The brittle deformation features in the Earth's upper crust are generally referred to as fractures. Fractures on which there has been significant shear displacement of rock are denoted faults. Fractures and faults are one of the dominating structures of the Earth's geology, as a large amount of the Earth's rock masses are fractured. Fractures and faults have a mechanical origin, and are the results of stress release in rock acted upon by tectonic forces, hydrostatic forces and gravity. The extent of fractures and faults can range from millimeters to thousands of kilometers. Examples of well known faults are the San Andreas fault in California and the faults in the East African rift valley. These are large scale faults and result from movements of the tectonic plates of the Earth's crust, see for instance Monroe and Wicander (1994). The impact of fractures at smaller scales was experienced during construction of the Romeriksporten tunnel on the railroad between Oslo, Norway, and the new international airport at Gardermoen. The building of the tunnel led to leakage into the tunnel of groundwater and surface water through the fracture network. The fact that Puttjern, a small, scenic lake in a recreation area, was heavily drained gave the problems great publicity.

In this work the focus is on faults, and the main issue is statistical analysis and stochastic modeling of faults and fault patterns in petroleum reservoirs. A brief introduction to geological faults is given below, presenting the terminology for faults used throughout this thesis. For a more thorough introduction to fractures and faults, and the properties of these geological structures, see for example Ramsay and Huber (1987) and Hatcher (1995). Motivations behind analysis and modeling of faults are presented, and a short summary of previous work on these topics is given.

Geological faults

When faulting occurs, movement of rock takes place along a surface, which is often steeply inclined, creating a discontinuity in the lithology referred to as the fault surface, see Figure 1:



(a) (b)

Figure 1: (a) Illustration of a fault network. Fault traces are observed on the surface on top. (b) Vertical cross section through a fault. Only rock inside a bounded volume is deformed or displaced.

Fault surface: A discontinuity in the rock, where blocks of rock on either side of the surface are displaced relative to each other. In general a fault surface is non-planar, as faults tend to have a rough surface, which in some cases can also be curved.

Fault plane: The plane in which the fault surface is embedded. For the purpose of mathematical fault modeling, a curved, rough fault surface can be approximately represented by a planar surface.

Figure 1 shows a network of fault surfaces, where blocks of rock inbetween are displaced and deformed due to faulting. The fault network can be interpreted as a collection of three dimensional single faults of finite extension. Each fault has a fault surface of limited extent where displacement of rock has occured along the fault surface. In addition, rock within an associated finite volume surrounding the fault surface has been moved relative to its original position and a continuous deformation of the blocks of rock has taken place. Generally the movement and deformation of rock are largest close to the center of the fault, defined as the center of the fault surface, and decreases continuously to zero in all directions away from the center, see Figure 1b.

Fault volume: Three dimensional volume of displaced and deformed rock.

Faults can be classified into three main groups: normal faults, reverse faults and strike-slip faults, which are illustrated in Figure 2a–c. A number of quantities can be used to describe a fault. Some of these are listed below, and are illustrated in Figure 2a for normal faults. The orientation of a fault is characterized by two quantities denoted strike and dip:

Strike direction: Orientation of intersection line between a fault plane and a horizontal plane.

Dip: Angle between a horizontal plane and a fault plane, measured perpendicular to the strike direction.

Dip direction: Orientation of a line within the fault plane oriented normal to the strike direction. The direction is taken towards increasing surface depth.

Normal faults and reverse faults are referred to as dip-slip faults, as the displacement of rock is parallel to the direction of the dip, while for strike-slip faults the displacement is in the direction of the strike. If rock is displaced both in the strike and dip direction, the fault is called an oblique-slip fault, see Figure 2d. Rock above the fault surface of a normal fault is denoted the hanging wall, and is displaced downwards relative to the rock below the fault surface, referred to as the foot wall, as illustrated in Figure 2a.

In geophysical literature, the term size of a fault is used both for the displacement, measuring the magnitude of movement of rock, and the volumetric extent of the fault. The displacement of a fault varies along the fault surface, and is defined as follows:

Displacement: The distance between two points on opposite sides of the fault surface that was infinitely close prior to faulting, measured in the fault plane.

Alternative quantities for representing the displacement exist:

Throw: Vertical component of the displacement.

Heave: Horizontal component of the displacement.

The term offset is also commonly used, although somewhat loosely, to describe displacement both quantitatively and qualitatively. In Part IV of this thesis, the term offset is used to describe the displacement observed when a fault plane intersects and creates a discontinuity in a seismic horizon.

In Part I of this thesis faults are represented as three dimensional objects embedded in a three dimensional space. The same nomenclature is used as in Walsh and Watterson (1988), where the extent of the fault surface is represented by:

Width: Maximum horizontal extent of the fault volume in the strike direction.

Length: Maximum extent of the fault volume in the dip direction.

When a three dimensional fault is observed in two dimensions, the fault surface appears as a single line, denoted a fault trace. A pattern of fault traces on a horizontal surface is illustrated in Figure 1a.

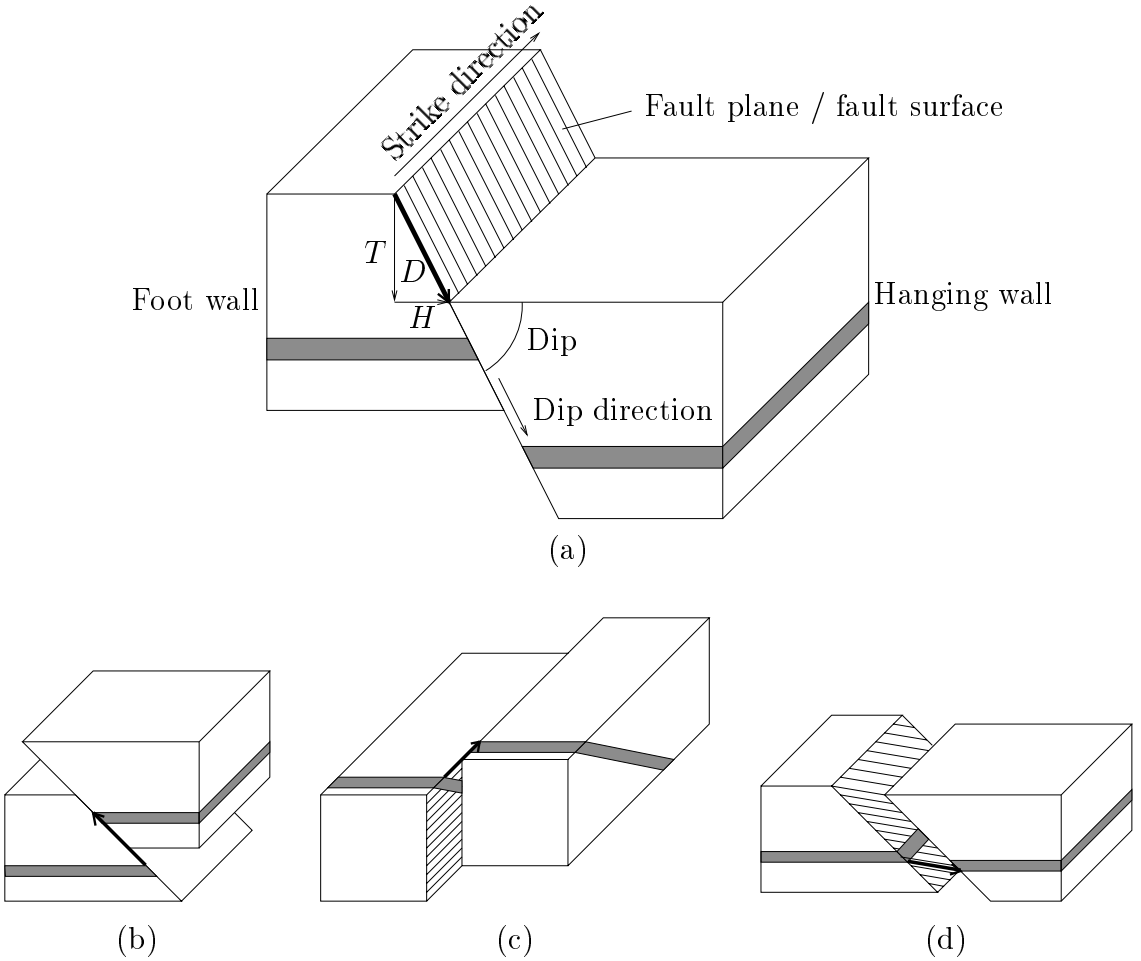


Figure 2: (a) Normal fault, (b) reverse fault, (c) strike-slip fault and (d) oblique-slip fault. The colored regions illustrate how a bed in a layered rock is offset by the fault. In figure (a) D is the displacement, T is the throw and H is the heave of the fault.

Fault trace: Line of intersection between a fault surface and an intersecting, two dimensional area of observations.

The two dimensional area of observations can for example be an outcrop surface or a seismic horizon. Parts III and IV of this thesis focus on fault traces in two dimensions, and in these papers the term length is used for the horizontal extent of a fault trace.

The definitions above apply to single faults. The following terms are used for the overall set of faults in the fault network, as illustrated in Figure 1a:

Fault population: Set of faults, where each fault is represented through displacement, strike, dip, width, length, etc.

Fault pattern: Spatial arrangement of a fault population.

Faults may also interact, and link up to form fault structures or arrays of faults. One type of fault array is en echelon faults, see Figure 3, which are in focus in Part III of this thesis.

En echelon faults: Structure of short, unconnected, overlapping fault segments with subparallel strikes.

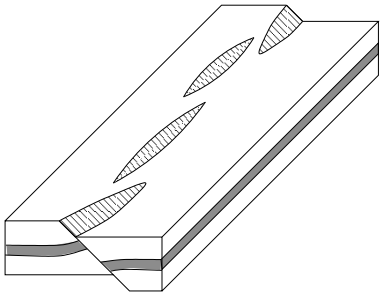


Figure 3: En echelon faults.

Fault analysis and modeling

There are several motives for the analysis and modeling of faults and fault patterns. The geologists improve their understanding of the geological nature of faults and the process of fault propagation, while the engineers use fault analysis and modeling as an integrated part in reservoir characterization. The latter is the focus of this study.

Analysis of fault populations found in literature can in general be divided into three categories: the statistical analysis of the distribution of fault size, represented as maximum displacement D or fault width W; the analysis of the relationship between the two quantities D and W; and the analysis of the spatial arrangement of faults. Cowie (1998) gives an overview with references on these topics. Over the last decades, fault and fracture analysis has mainly been performed under the assumption of a fractal nature of faulting, see King (1983), involving fractal distributions and fractal scaling relationships. Fractal distributions are also termed power law distributions.

In the petroleum industry, production forecasts are based on reservoir characterization. Knowledge about fractures and faults are of great importance in the modeling of petroleum reservoirs, as they have influence on the flow properties of the rock and compartmentalization of the reservoir, and thus on the oil and gas recovery, see for instance Omre et al. (1994), Yielding et al. (1997), Jones et al. (1998) and Manzocchi et al. (1999). Open fractures constitute channels for fluid flow, while sealing fractures or faults can act as barriers to flow. The overall flow in open fractures depends on the connectivity of the fracture network, which again depends on the fracture intensity and on the distribution of fracture sizes. Faults can disconnect layers of rock, influencing the connectivity of the hydrocarbon bearing lithologies, and can give rise to oil and gas traps.

The presence of faults is crucial in the modeling of groundwater flow. For instance, flow properties of the groundwater in fractured and faulted rocks are of concern when planning radioactive waste deposits. Analysis of faults and fault patterns also has a long tradition in the mining industry. Furthermore, mapping of faults is of importance in the construction of infrastructure, for example in the building of roads, bridges, tunnels, dams, etc.

Although faults are geological structures embedded in three dimensions, observations of faults are usually limited to only one or two dimensions, and exact three dimensional imaging of sub-surface faults is in general not obtainable. Faults at microscopic scales can be observed in core samples and bore holes, giving accurate observations within a strictly limited volume. Faults at macroscopic scales can be observed in one or two dimensions in wells, coal mines, outcrops or through seismic surveys. Interpretation of seismic observations can also give indications of the three dimensional pattern of faults at large scales. Some well observations are typically available for petroleum reservoirs, but the major source of information comes from seismic data. However, only faults of displacement above $10 - 30m$ are easily resolved from the seismic data, while the typically greater number of sub-seismic faults are unobservable.

Fault size distribution and fault numbers

Fault size distributions are the topic of Part I and Part III of this thesis. Fault sizes, represented as either maximum displacement or width, are frequently assumed to follow a

power law, or fractal, distribution. This assumption is supported by numerous field studies, see for instance JSG (1996, Vol. 18). According to this power law, a fault population should satisfy

$$P(X \geq x) = \text{const} \times x^{-\beta}, \tag{1}$$

where X is the fault size and $\beta > 0$ is denoted the power law exponent. The probability (1) can be estimated by the relative number $\hat{P}(X \geq x) = N_x/N_T$, where N_x is the number of faults larger than x and N_T is the total number of observations. Statistical analysis of fault size distributions is then concentrated on estimating the parameter β. Inference is frequently based on the expression

$$\log N_x = \log N_T + \log \hat{P}(X \geq x) = \hat{\alpha} - \hat{\beta} \log x,$$

which is often evaluated by visual inspection. A least square linear regression is performed on log-log scale using the sorted observations $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N_T)}$ as explanatory variables and $N_T, N_T - 1, \ldots, 2, 1$ as response variables. This gives an estimator for β, however with mean and variance unknown. Alternative estimators, for example maximum likelihood estimators, are discussed by Pickering et al. (1995) and Clark et al. (1999). Estimation techniques accounting for sampling biases are also suggested, see Walsh et al. (1994) and Pickering et al. (1996).

Although the belief in a power law distribution of fault and fracture size has proved persistent, there are studies where other distributions are found to be more adequate, for example an exponential distribution, see Cowie et al. (1994), Spryopoulos et al. (1999) and Ackermann et al. (1999), and a log-normal distribution, see Odling (1997).

In reservoir modeling the aim of the statistical fault analysis is to predict the number of faults of different sizes below seismic resolution. By assuming faults, including both seismic and sub-seismic scales, having a common size distribution, the expected proportion of large faults can be determined and the overall number of faults can be estimated. Modeling of sub-seismic size distributions is based on extrapolation of the size distribution observed at seismic scales, possibly supported by well observations, see Badley et al. (1990), Childs et al. (1990), Gauthier and Lake (1993). Furthermore, extrapolation of the distribution from the dimension of observations to the three dimensional reservoir is studied, see Heffer and Bevan (1990), Marrett and Allmendinger (1991), Piggott (1997) and Berkowitz and Adler (1998). When the results of an analysis of faults is used for extrapolation into unobserved scales or dimensions, it is of great importance that the extrapolation is performed under an appropriate model. If the model does not reflect the true distribution of fault sizes, the number of faults estimated through extrapolation is unreliable. The uncertainty in the estimated number of faults contains uncertainty both in the model assumption and in the estimates of the corresponding model parameters.

Fault patterns

Modeling of fault and fracture networks can be made either by modeling of the physical process of faulting or by descriptive modeling of existing fault patterns. Numerical modeling, see for instance Cowie et al. (1993) and Malthe-Sørenssen et al. (1998), and experimental modeling, see for example Sornette et al. (1993), McClay (1996) and Spryopoulos et al. (1999), are applied to study the geological process of fault growth. These techniques are thus aimed at gaining further knowledge about the physical processes that control faulting. Descriptive modeling, providing a static representation of a fault pattern, is made through spatial stochastic modeling of faults.

Stochastic modeling of the fault pattern of a petroleum reservoir can be used as an integrated part in stochastic reservoir characterization. The uncertainties in the production forecast of a reservoir can be studied through repeated sampling of the reservoir properties, followed by running a fluid flow simulator for each realization, see Lia et al. (1997). Faults above seismic resolution can be interpreted from data, while stochastic models are mainly created aiming at modeling sub-seismic faults and fractures, see for example Chilès (1988), Brand and Haldorsen (1988), Gauthier and Lake (1993) and Munthe et al. (1993). The results of a fault analysis is often used as input to the stochastic modeling, where conditioning on data is performed in the sense that spatial fault intensity and distributions of fault width, displacement, strike, dip, etc. are based on analysis of available data.

Thesis Summary

When fault size distributions are extrapolated beyond the scale and dimension of observations, the choice of model becomes very important. Model choice for fault size distribution is frequently based on visual interpretations of the empirical cumulative distribution functions. Part I of this thesis presents a statistical approach to model choice. To the knowledge of the author, statistical tools for model choice have previously not been used for fault size distributions. Fault size is represented by either maximum displacement or fault width, and Part I compares the frequently used power law distribution with the exponential distribution. The model choice problem is defined in a Bayesian framework and Bayes factor is used as a criterion for model choice, see Kass and Raftery (1995). The relationship between the size distribution in three dimensions and the corresponding distribution in the dimension of observations is considered, since observations from the lower dimension are used to choose a model for the fault population in three dimensions. Furthermore, sampling errors like missing observations, measurement errors, censoring, bias, etc. are accounted for. Prior distributions are defined for the parameters of the competing models and the number of faults, while sampling errors are modeled by the likelihood function. The paper presents a sensitivity analysis for a seismic dataset from the Gullfaks field in the North Sea, exploring the influence of the prior and likelihood parameters on the conclusion of the

model choice. Well observations from the same field are used to calibrate the prior mean number of faults. An early version of this work was presented at the conference Modelling Permeable Rocks, Cambridge, UK, March 1998, see Appendix A.

Bayes factors are in general not analytically tractable, and this is also the case in the model choice problem presented in Part I. Markov chain Monte Carlo (McMC) sampling techniques can be used to estimate Bayes factor, see Han and Carlin (2000) for an overview, and Part II of the thesis presents an McMC algorithm constructed for the type of problem presented in Part I. Under both of the competing models, the initially high dimensional posterior probability density functions (pdf) can be reduced to pdfs of low dimensional variables through integration of analytically tractable components of the unknown variables. The target distribution of the McMC algorithm is a joint distribution of a model indicator and parameters under both competing models, and Bayes factor and the model parameters can be estimated based on samples from the algorithm. The joint pdf is defined through the introduction of pseudopriors, see Carlin and Chib (1995). There is a freedom of choice when defining the pseudopriors, and this freedom is explored to find optimal choices with respect to uncertainty in the estimation of Bayes factor. The performance of the McMC algorithm is studied for the case of no sampling errors, in which case Bayes factor is analytically tractable and the estimates can be compared with the true values.

Published results on the relationship between fault size distributions in different dimensions are all based on the assumption of independently, uniformly distributed fault positions, and are in general based on power law size distributions. However, natural datasets are not always found to obey the theoretically obtained relationship. Part III of this thesis discusses possible sources of deviations, presenting analytical considerations of problems encountered by the geophysicists. The paper focuses on the influence of spatial clustering of faults and interacting faults linked into fault arrays. Furthermore, the paper illustrates possible consequences if extrapolation of fault size distributions into higher dimensions are performed based on a wrong assumption of a power law distribution. The paper is written for a geophysical audience and the terminology is mainly chosen according to geophysical tradition, hence deviating somewhat from the statistical terminology used elsewhere in the thesis. The collaborative work of Part III was initiated while I was visiting the Department of Geology and Geophysics, the University of Edinburgh, UK.

Part IV presents a spatial stochastic model for fault patterns at seismic scales. Faults at this scale are traditionally mapped by human interpreters, while recently algorithms for tracking faults have been developed, for example the coherence cube presented by Bahorich and Farmer (1995). To the knowledge of the author, Part IV of this thesis presents for the first time a stochastic approach to fault detection at seismic scales. The model is defined in a Bayesian framework, and includes conditioning to three dimensional seismic data and well observations. Through stochastic modeling, assessment of uncertainty in fault patterns at seismic scales can be obtained. The prior distribution is used to model a faulted horizon, including fault traces and displacement profiles along the traces, according to geological

expectations. The faulted horizon is represented by a hexagonal grid image, where the prior pdf of fault traces is based on the work by Tjelmeland and Besag (1998). The relationship between the faulted horizon, three dimensional seismic data and well observations is defined through the likelihood function. The likelihood function is based on a Gaussian model for reflection coefficients of a non-faulted reservoir, where reflection coefficients are linked with seismic data through a linear convolution model, see Eide (1999). Based on this model, the corresponding distribution of the observations of a faulted reservoir is obtained. An McMC algorithm is constructed to sample from the posterior distribution of fault patterns, where the implementation of the algorithm is a further development of the implementation used by Tjelmeland and Besag (1998). The model and the algorithm works well for a synthetic dataset. A real seismic dataset is also considered, but the results are not found to be satisfactory. Problems with applying the model and the algorithm to these real data are discussed. This work was presented at the conference EAGE/SPE international symposium on petroleum geostatistics, Toulouse, France, April 1999 and at the 6th International Geostatistics Congress, Cape Town, South Africa, April 2000, see Appendix B and C.

Part V originates from a discussion between my supervisor Henning Omre and Dr. John Walsh, the Fault Analysis Group, University of Liverpool, UK. The discussion concerned different schemes for partitioning of a line segment, and the corresponding length distributions of the resulting shorter segments. Some possible partitioning schemes are considered in Part V, and the analytically obtained length distributions of the shorter segments are discussed. The topic of Part V is related to analysis and modeling of faults or fractures observed in one dimension, where fault positions constitute partitioning points along a line and fault separation lengths are represented by the shorter segments, see Gillespie et al. (1993).

References

Ackermann, R. V., Withjack, M. O. and Schlische, R. W. (1999). The geometric and statistical evolution of normal fault systems: An experimental study of the effects of mechanical layer thickness on scaling laws, submitted to *Journal of Structural Geology*.

Badley, M. E., Freeman, B., Roberts, A. M., Thatcher, J. S., Walsh, J., Watterson, J. and Yielding, G. (1990). Fault Interpretation During Seismic Interpretation and Reservoir Evaluation, *The Integration of Geology, Geophysics, Petrophysics and Petroleum Engineering in Reservoir Delineation, Description and Management*, Proceedings of the 1st Archie Conference, Houston, Texas, pp. 224–241. AAPG.

Bahorich, M. S. and Farmer, S. L. (1995). The coherence cube, *The Leading Edge* **14**(10): 1053–1058.

Berkowitz, B. and Adler, P. M. (1998). Stereological analysis of fracture network structure in geological formations, *Journal of Geophysical Research* **103**(B7): 15339–15360.

Brand, P. J. and Haldorsen, H. H. (1988). Modelling large-scale heterogeneities caused by faulting with a stochastic approach, *Revue de l'Institut français du pétrole* **43**(5): 647–657.

Carlin, B. P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods, *J. R. Statist. Soc. B* **57**(3): 473–484.

Childs, C., Walsh, J. J. and Watterson, J. (1990). A Method for Estimation of the Density of Fault Displacements below the Limits of Seismic Resolution in Reservoir Formations, *North Sea Oil and Gas Reservoirs - II*, Graham & Trotman, pp. 309–318.

Chilès, J. P. (1988). Fractal and Geostatistical Methods for Modeling of a Fracture Network, *Mathematical Geology* **20**(6): 631–654.

Clark, R. M., Cox, S. J. D. and Laslett, G. M. (1999). Generalization of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats, *Geophysical Journal International* **136**: 357–372.

Cowie, P. A. (1998). Normal Fault Growth in Three-Dimensions in Continental and Oceanic Crust, *Faulting and Magmatism at Mid-Ocean Ridges*, Geophysical Monograph 106, the American Geophysical Union, pp. 325–348.

Cowie, P. A., Malinverno, A., Ryan, W. B. F. and Edwards, M. H. (1994). Quantitative fault studies on the East Pacific Rise: A comparison of sonar image techniques, *Journal of Geophysical Research* **99**(B8): 15205–15218.

Cowie, P. A., Vanneste, C. and Sornette, D. (1993). Statistical Physics Model for the Spatiotemporal Evolution of Faults, *Journal of Geophysical Research* **98**(B12): 21809–21821.

Eide, A. L. (1999). *Stochastic Reservoir Characterization Constrained by Seismic Data*, PhD thesis, Norwegian University of Science and Technology. Doktoringeniøravhandling 1999:5.

Gauthier, B. D. M. and Lake, S. D. (1993). Probabilistic Modeling of Faults Below the Limit of Seismic Resolution in Pelican Field, North Sea, Offshore United Kingdom, *The American Association of Petroleum Geologists Bulletin* **77**(5): 761–777.

Gillespie, P. A., Howard, C. B., Walsh, J. J. and Watterson, J. (1993). Measurement and characterisation of spatial distributions of fractures, *Tectonophysics* **226**(1-4): 113–141.

Han, C. and Carlin, B. P. (2000). MCMC Methods for Computing Bayes Factors: A Comparative Review, *Research Report 2000-001*, Division of Biostatistics, University of Minnesota.

Hatcher, R. D., Jr. (1995). *Structural Geology. Principles, Concepts, and Problems*, 2 edn, Prentice Hall, Englewood Cliffs, New Jersey.

Heffer, K. J. and Bevan, T. G. (1990). Scaling Relationships in Natural Fractures: Data, Theory, and Application, *Europec 90*, Society of Petroleum Engineers, pp. 367–376. SPE Paper 20981.

Jones, G., Fisher, Q. J. and Knipe, R. J. (eds) (1998). *Faulting, Fault Sealing and Fluid Flow in Hydrocarbon Reservoirs*, number 147 in *Special Publications*, Geological Society, London.

JSG (1996). *Journal of Structural Geology*, Vol. 18, No. 2/3, Special Issue, Scaling Laws for Fault and Fracture Populations – Analyses and Applications.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association* **90**(430): 773–795.

King, G. (1983). The Accommodation of Large Strains in the Upper Lithosphere of the Earth and Other Solids by Self-similar Fault Systems: the Geometrical Origin of b-Value, *Pure and Applied Geophysics* **121**(5/6): 761–815.

Lia, O., Omre, H., Tjelmeland, H., Holden, L. and Egeland, T. (1997). Uncertainties in Reservoir Production Forecasts, *AAPG Bulletin* **81**(5): 775–802.

Malthe-Sørenssen, A., Walman, T., Feder, J., Jøssang, T., Meakin, P. and Hardy, H. H. (1998). Simulation of extensional clay fractures, *Physical Review E* **58**(5): 5548–5564.

Manzocchi, T., Walsh, J. J., Nell, P. and Yielding, G. (1999). Fault transmissibility multipliers for flow simulation models, *Petroleum Geoscience* **5**(1): 53–63.

Marrett, R. and Allmendinger, R. W. (1991). Estimates of strain due to brittle faulting: sampling of fault populations, *Journal of Structural Geology* **13**(6): 735–738.

McClay, K. R. (1996). Recent advances in analogue modelling: uses in section interpretation and validation, *in* P. C. Buchanan and D. A. Nieuwland (eds), *Modern Developments in Structural Interpretation, Validation and Modeling*, Geological Society Special Publication No. 99, pp. 201–225.

Monroe, J. S. and Wicander, R. (1994). *The Changing Earth*, West Publishing Company.

Munthe, K. L., Omre, H., Holden, L., Damsleth, E., Heffer, K., Olsen, T. S. and Watterson, J. (1993). Subseismic Faults in Reservoir Description and Simulation, *68th Annual Conference and Exhibition of the Society of Petroleum Engineers*, Society of Petroleum Engineers, pp. 843–850. SPE Paper 26500.

Odling, N. E. (1997). Scaling and connectivity of joint systems in sandstones from western norway, *Journal of Structural Geology* **19**(10): 1257–1271.

Omre, H., Sølna, K., Dahl, N. and Tørudbakken, B. (1994). Impact of Fault Heterogeneity in Fault Zones on Fluid Flow, *North Sea Oil and Gas Reservoirs - III*, Norwegian Institute of Technology, Kluwer Academic Publishers, pp. 185–200.

Pickering, G., Bull, J. M. and Sanderson, D. J. (1995). Sampling power-law distributions, *Tectonophysics* **248**: 1–20.

Pickering, G., Bull, J. M. and Sanderson, D. J. (1996). Scaling of fault displacements and implications for the estimation of sub-seismic strain, *in* P. C. Buchanan and D. A. Nieuwland (eds), *Modern Developments in Structural Interpretation, Validation and Modeling*, Geological Society Special Publication No. 99, pp. 11–26.

Piggott, A. R. (1997). Fractal relations for the diameter and trace length of disc-shaped fractures, *Journal of Geophysical Research* **102**(B8): 18121–18125.

Ramsay, J. G. and Huber, M. I. (1987). *The Techniques of Modern Structural Geology*, Vol. 2: Folds and Fractures, Academic Press.

Sornette, A., Davy, P. and Sornette, D. (1993). Fault Growth in Brittle-Ductile Experiments and the Mechanics of Continental Collisions, *Journal of Geophysical Research* **98**(B7): 12111–12139.

Spryopoulos, C., Griffith, W. J., Scholz, C. H. and Shaw, B. E. (1999). Experimental evidence for different strain regimes of crack populations in a clay model, *Geophysical Research Letters* **26**(8): 1081–1084.

Tjelmeland, H. and Besag, J. (1998). Markov Random Fields with Higher-Order Interactions, *Scandinavian Journal of Statistics* **25**(3): 415–433.

Walsh, J. J. and Watterson, J. (1988). Analysis of the relationship between displacements and dimensions of faults, *Journal of Structural Geology* **10**(3): 239–247.

Walsh, J. J., Watterson, J. and Yielding, G. (1994). Determination and Interpretation of Fault Size Populations: Procedures and Problems, *North Sea Oil and Gas Reservoirs - III*, Norwegian Institute of Technology, Kluwer Academic Publishers, pp. 141–155.

Yielding, G., Freeman, B. and Needham, D. T. (1997). Quantitative fault seal prediction, *The American Association of Petroleum Geologists Bulletin* **81**(6): 897–917.

Part I

Model choice for fault size distribution

Model choice for fault size distribution

Hilde G. Borgos

Abstract

Information about the number of faults and the fault size distribution is an important factor in reservoir modeling and fluid flow simulations. Inference is often made based on an assumption of a negative power law distribution of maximum displacements or fault widths. In this work it is shown how this assumption can be statistically validated, by comparing the negative power law distribution with an exponential distribution using Bayes factor. Inference about three dimensional fault size distributions are often based on observations obtained in lower dimension. Thus the relationship between the distribution in three and two dimensions must be known, for both the negative power law and the exponential distribution. Furthermore, observation errors, like bias, censoring and missing observations, are accounted for in the model choice problem.

1 Introduction

When tectonic forces, forces from overburdens, hydrostatic pressure etc. act on a subsurface rock, stress fields arise. Some of the stress can be released through the formation of faults and subsequent displacement of rock. A fracture in the rock forms, and if shear forces are present the rock in a volume on both sides of the fracture is displaced, creating a fault. See for instance Hatcher (1995) for an introduction to faults. The fault geometry is defined by the finite extension of the faulted volume. The displacement is typically at its maximum near the center of the faulted volume, descending to zero at the boundaries of the fault. Figure 1 gives a sketch of a vertical cross section through a fault, where the lines illustrate the displacement of rock.

When a stress field is present in a rock, a number of faults can be produced, resulting in a pattern of faults. The number of faults and the location, geometry and displacement of each fault in the fault pattern can have great influence on fluid flow in a reservoir. Faults can act as either channels or barriers to fluid flow, and compartmentalizing of reservoirs can be related to the fault pattern, see for example Yielding et al. (1997) and Jones et al. (1998). Thus the effect of faults should be taken into account in reservoir modeling and

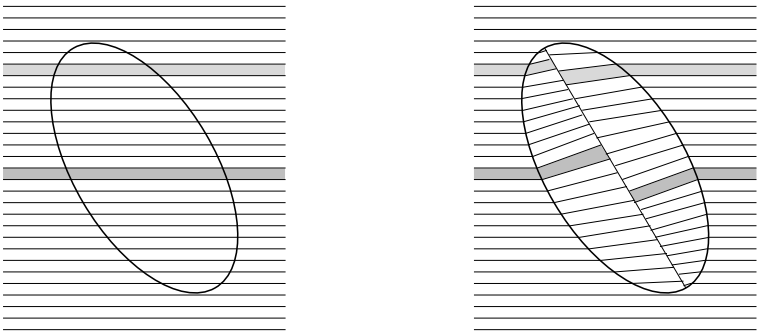


Figure 1: Vertical cross section through a fault. To the left is an illustration of the rock prior to faulting, with the fault geometry indicated. The right figure shows the rock after faulting. Only rock inside a bounded volume is displaced.

fluid flow simulations. However, the fault pattern of a subsurface reservoir can not be observed exactly, but inference can be made based on available observations. For example, it is of interest to predict the number of faults in a reservoir and their size distribution.

Observations of faults often span only a few size scales. Seismic data include faults on a large scale, while core samples provide observations on a small scale and within a limited area. Extrapolation of fault size distributions obtained at seismic scales, combined with core observations if available, can be used to predict the distribution of fault sizes and the number of faults below seismic resolution, see for instance Badley et al. (1990), Childs et al. (1990), Walsh et al. (1991, 1994), Sassi et al. (1992), Gauthier and Lake (1993). Several uncertainties are related to the extrapolation of fault distributions, as discussed in the cited papers. For example, errors in the fault sampling may lead to great uncertainties in estimated model parameters. Furthermore, changes in geological properties beyond the scales of observations can not be predicted based on the observations, thus extrapolation of the estimated distribution can lead to unreliable results.

In this work the focus is on the size distribution of faults, where fault size is measured as either maximum displacement or width. The width represents the maximum horizontal extent of the fault. In order to make inference about the overall fault population based on available observations, an appropriate model should be chosen to represent the distribution of fault size, and the model parameters should be estimated. Model choice is important for predictive purposes. However, the model choice can also influence on the understanding of geological aspects related to the process of faulting.

Faults are objects embedded in a three dimensional space, while observations are often obtained in a lower dimension. Two dimensional observations are obtained from seismic maps or from surface outcrops of the rock, while one dimensional observations consist of measurements along a line. The size distribution observed in a lower dimension does not necessarily equal the original distribution, since large faults have a greater probability than

small ones of being observed in lower dimensions and may thus be over-represented, see for example Heffer and Bevan (1990), Marrett and Allmendinger (1991). Hence, inference about the size distribution of faults in three dimensions can only be made if the relationship to the observations in lower dimensions is known. Another problem encountered when investigating the fault size distribution is observation errors, which should also be accounted for when choosing a proper model for the distribution.

If a fault size distribution is defined in three dimensions, corresponding unique distributions exist both in intersecting two dimensional planes and one dimensional lines. Under reasonable assumptions on fault geometry and spatial distribution of faults, these distributions can be obtained by simulations or forward calculations. Examples of forward calculations are found in the literature, mainly in the case of negative power law, or fractal, distributions of fault size. Heffer and Bevan (1990), Piggott (1997) and Berkowitz and Adler (1998) studied relationships between fault or fracture populations in three and two dimensions, assuming faults in three dimensions are shaped as circular discs. Marrett and Allmendinger (1991) presented a forward calculation performed without defining a specific model for the fault geometry, but using a general relationship $A \propto W^2$ where A is the surface area of the fault and W the fault width. A basic assumption in all the cited papers is that fault positions are independently, uniformly distributed in space.

Consider a hypothesis testing problem concerning fault size distributions in three dimensions. Inference can be made based on distributions in lower dimensions derived by forward calculations, and observations obtained in two or one dimensions can be used to decide which of two proposed distributions are most likely to describe the fault sizes in three dimensions. The conclusion only tells which of the two models is favored based on the observations, while there is still a possibility that none of the two distributions is the right one. For example, for a size distribution obtained in lower dimensions, the initial distribution in three dimensions need not be unique.

To the knowledge of the author, this work presents for the first time a statistical approach to dealing with ambiguities related to choosing a proper fault size distribution. Two possible distributions of maximum displacement of faults in three dimensions are considered: a fractal distribution and an exponential distribution. Corresponding distributions of fault width are obtained, and distributions of fault sizes observed in lower dimensions are derived under both models. A model choice problem is defined, comparing the two possible models for observations gathered in two dimensions. The uncertainty in the model choice is closely related to sampling errors, which are accounted for in the stochastic model. The conclusion of the model choice problem is based on Bayes factor, see Kass and Raftery (1995) for an overview. Bayes factor can not be analytically obtained for the stochastic model presented in this work, but is estimated using the Markov chain Monte Carlo (McMC) algorithm presented by Borgos (2000). Only if all observation errors are ignored can Bayes factor be calculated analytically.

The report is organized as follows. In Section 2 the fault model used in this work is

presented. In Section 3 it is described how the distribution of fault size in lower dimensions can be derived from the distribution in three dimensions. This procedure is applied in Section 4, for both the fractal and the exponential model. Fault observations are discussed in Section 5, and the distribution of fault sizes and the number of faults, conditioned to the observations, is defined in a Bayesian framework. In Section 6 it is described how Bayes factor can be used to compare the two suggested models, and examples are provided in Sections 7 and 8.

2 Fault Representation

A single fault is a three dimensional object, and can be described through a number of fault characteristics. Throughout this work, fault geometry denotes the shape and extent of the faulted volume. Figure 1 shows a vertical cross section through a fault. A fault plane is a plane through the center of a fault, and the intersection of the faulted volume and the fault plane is called the fault surface. The fault surface forms a discontinuity in the rock, where rock on one side of the surface is displaced relative to the rock on the opposite side. Away from the fault surface, rock within the faulted volume is still continuous, and is only displaced relative to its original position. Consider two points infinitely close prior to faulting, but located on opposite sides of the fault surface. The distance between the two points after faulting, measured in the fault plane, is denoted the displacement. The displacement is typically at its maximum near the center of the fault surface, decreasing to zero along the fault surface and within the displaced volume (Walsh and Watterson, 1987). The angle between the fault plane and the horizontal plane is called the dip of the fault, and the direction of the line of intersection between the fault plane and the horizontal plane is commonly known as the strike. The strike is measured as the angle between the line and a reference direction, for example north.

In this study an individual fault is represented by its location, orientation, the fault geometry and the displacement within the faulted volume. The location of a fault is defined at its center point. To enable fault modeling, the complexity of fault geometries observed in nature is simplified, and the geometry is represented by a parametric model. The parameters include the volumetric extent and orientation of the fault. A simplified parametric model is also used for the displacement within the faulted volume.

2.1 Fault model

A fault pattern can be modeled as a marked point process where the points are the fault locations and the marks are sets of fault characteristics, see Chilès (1988), Munthe et al. (1993), Munthe et al. (1994), Stoyan and Stoyan (1994) and Stoyan et al. (1995). The marked point process is given as follows:

Location The location of a fault is defined as its center point, and is represented by the stochastic variable $\boldsymbol{\xi}$. In three dimensions the location is given by $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$.

Geometry The fault geometry $G = G(\Psi)$ is a parametric model representing the shape and extent of the faulted volume.

Parameterization The fault geometry is parameterized by a set of stochastic variables denoted Ψ. The parameterization includes the extents in three orthogonal directions: the horizontal extent W, the extent L in the dip direction and the extent R perpendicular to the fault plane. The angles of dip Θ and the strike Φ is also included, thus $\Psi = (W, L, R, \Theta, \Phi)$.

Maximum displacement The maximum displacement of the fault, obtained at the center point $\boldsymbol{\xi}$, is given by a stochastic variable D.

Displacement function For a fault with maximum displacement $D = d$, the displacement at any point u within the faulted volume is given by the displacement function $\delta(u; d)$. The function should satisfy the following conditions:

$$\begin{array}{lll} \text{(i)} & \delta(u;d) = d, & u = \boldsymbol{\xi} \\ \text{(ii)} & \delta(u;d) \leq d, & u \text{ inside } G(\Psi). \\ \text{(iii)} & \delta(u;d) \equiv 0, & u \text{ at the boundary or outside of } G(\Psi). \end{array}$$

Fault representation One fault can now be represented as a stochastic variable M, given as $M = (\boldsymbol{\xi}, \Psi, D, G(\Psi), \delta(\cdot))$, constituting a marked point in the marked point process.

In literature, the term size is used to represent both the width of a fault and the maximum displacement. In the first case size is a measure of the horizontal extent of the fault. In the latter case the maximum relative movement on opposite sides of the fault surface is measured. Fault width and maximum displacement are related, see Watterson (1986), and throughout this work the word size is used as a general expression for either quantity.

2.2 Fault geometry

The fault geometry $G(\Psi)$ in three dimensions is represented by an ellipsoid (Walsh and Watterson, 1988, 1989), as illustrated in Figure 2a. A vertical cross section of the fault ellipsoid is shown in Figure 2b, where the intersection of the ellipsoid with a plane results in an ellipse. The ellipsoid is parameterized by the axis lengths $W/2$, $L/2$ and $R/2$ respectively, measured from the center point to the boundary of the ellipsoid, and by the dip Θ and strike Φ, see Figure 2. The strike is of no concern throughout this work, and all faults

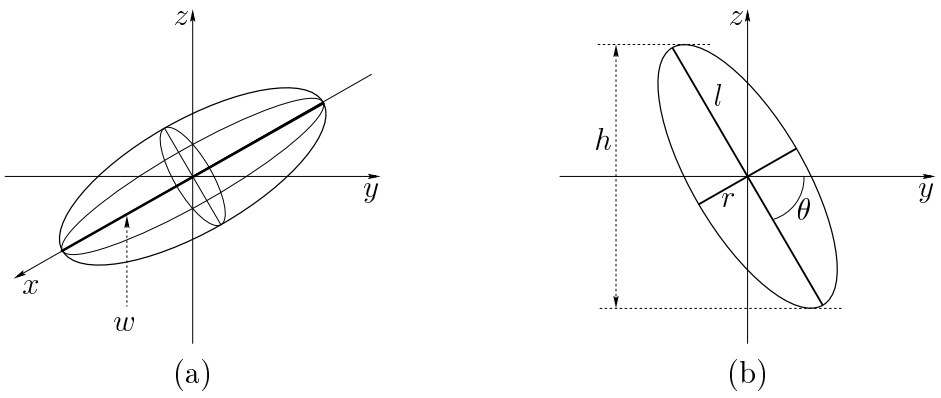


Figure 2: (a) Fault ellipsoid in three dimensions. The length of the horizontal major axis is $w/2$, measured from the center point to the ellipsoid boundary. The lengths of the two other axes are $l/2$ and $r/2$, and are illustrated in figure (b), which shows a vertical cross section of the fault ellipsoid perpendicular to fault strike. The dip of the fault is θ and the vertical extent is h.

are assigned equal strike $\Phi = 0$ without loss of generality. For a fault located at the origin, the geometry $G(\psi)$ is then the ellipsoid described by the set of points in $\mathbb{R}^3$:

$$G(\psi) = G(w, l, r, \theta, 0) = \left\{ (x, y, z)\,; \frac{x^2}{w^2} + \frac{(y \sin\theta - z\cos\theta)^2}{l^2} + \frac{(y\cos\theta + z\sin\theta)^2}{r^2} = \frac{1}{4} \right\}.$$

When faulting occurs, only the volume inside of the ellipsoid is deformed by the fault, while the volume outside the ellipsoid is unaltered. The displacement inside the fault ellipsoid is given by the maximum displacement D and the displacement function $\delta(\cdot)$. Deterministic relationships between D, W, L and R are assumed:

$$W = (Dc_1)^{1/\tau}, \quad L = Wc_2^{-1}, \quad R = c_3, \tag{1}$$

where c_1, c_2 and c_3 and τ are constants, see Watterson (1986), Walsh and Watterson (1987, 1988, 1989), Cowie and Scholz (1992a, b), Gauthier and Lake (1993), Dawers et al. (1993). Several case studies have found values of the parameter τ in the range $[1, 2]$, and although the relationship $W \propto D^{1/\tau}$ is widely accepted, there has been some disagreement about the actual value of τ. See Cowie (1998) for a discussion and for further references.

The vertical extent, or height, of the ellipsoid is given by a function $h(\cdot)$ depending on the lengths of the axes and the dip, and is illustrated in Figure 2b. Since there is a deterministic relationship between D, W, L and R, the height can be expressed as a function of either of these, for example

$$h(w, \theta) = \left((wc_2^{-1})^2 \sin^2\theta + c_3^2\cos^2\theta\right)^{1/2} \approx wc_2^{-1}\sin\theta \tag{2}$$

$$h(d, \theta) = \left(((dc_1)^{1/\tau}c_2^{-1})^2 \sin^2\theta + c_3^2\cos^2\theta\right)^{1/2} \approx (dc_1)^{1/\tau}c_2^{-1}\sin\theta. \tag{3}$$

The approximations of the height function correspond to the height of the axis of the fault ellipsoid in the dip direction. The length of the ellipsoid in the dip direction is L and the

dip is Θ, resulting in a vertical extent of $L \sin \Theta$. Combining this with the expressions in (1) gives the approximations in Expressions (2) and (3). The approximations of the heights are used in later sections due to their simple analytical forms. A discussion of the approximations and the error terms is given in Appendix A.

Consider a fault of width $W = w$, measured through the center of the ellipsoid. At a vertical distance $uh(w, \theta)/2$ from the center, where $u \in [-1, 1]$, the horizontal extent of the ellipsoid is

$$w'(u; w) = w\sqrt{1 - u^2}, \tag{4}$$

see Figure 3. Thus, if a horizontal plane intersects a fault of width $W = w$ at a vertical distance $uh(w, \theta)/2$ from the center of the fault, the width of the fault ellipse in this plane is $w'(u; w)$.

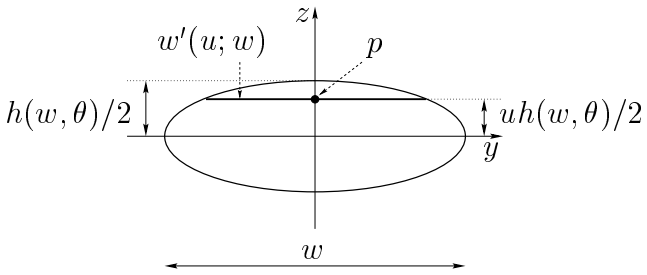


Figure 3: Vertical projection of the fault ellipsoid parallel to fault strike. At a vertical distance $uh(w, \theta)/2$ from the center of the fault ellipsoid, the horizontal extent is given by $w'(u; w)$ in Expression (4). At the point p the displacement is given by $\delta(u; d)$ in Expression (5).

2.3 Displacement function

The displacement function should satisfy the conditions (i)-(iii) given in the beginning of Section 2. In this work, a general displacement function of the form

$$\delta(u; d) = (1 - |u|)^{\gamma} d, \qquad -1 \leq u \leq 1, \ \gamma > 0, \tag{5}$$

is used. The parameter γ controls the shape of the displacement function along the axis in the dip direction, and some examples are illustrated in Figure 4. Throughout this work, γ is assumed to be a constant. See Walsh and Watterson (1987), Barnett et al. (1987), Dawers et al. (1993), and Dawers and Anders (1995) for discussions on displacement functions. If a horizontal plane intersects a fault with maximum displacement $D = d$ at a vertical distance $uh(d, \theta)/2$ from the center of the fault, see Figure 3, the maximum displacement of the fault in this plane is $\delta(u; d)$.

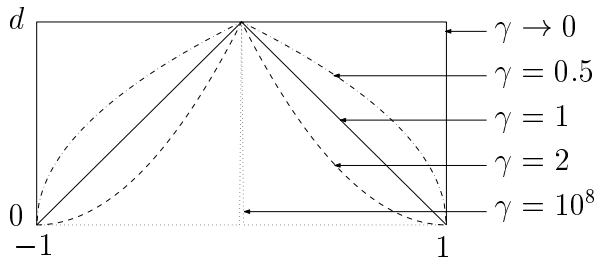


Figure 4: The displacement function $\delta(u; d)$ in Expression (5), with different values of γ.

3 Relationship Between Fault Distributions in One, Two and Three Dimensions

If a stochastic model for the fault geometry is well specified and a distribution of fault size in three dimensions is defined, the distribution in two dimensions can be found by forward calculations. The calculations may lead to expressions that are not analytically tractable, in which case numerical approximations can be used.

Consider faults spatially distributed within a three dimensional box $\mathcal{B}$ of vertical extent T_v and horizontal extent $T_{h1} \times T_{h2}$. The box is given by the set $\mathcal{B} = [0, T_v] \times [0, T_{h1}] \times [0, T_{h2}]$. If a two dimensional horizontal plane intersects $\mathcal{B}$ at an arbitrary depth $t \in [0, T_v]$, a number of faults inside $\mathcal{B}$ are intersected by the plane. The geometry of the intersection between the plane and an individual fault is a two dimensional ellipse, and within this intersection area the displacement is at a maximum at the center point of the ellipse. The stochastic variable X is now used as a common term for the size of a fault measured in three dimensions, being either maximum displacement or width. The corresponding size measured in two dimensions is denoted Y. The probability density functions (pdf) of the variables are denoted $f(x)$ and $f(y)$ respectively. The pdf $f(x)$ in three dimensions is assumed to be known, and the aim is to derive the corresponding pdf $f(y)$ in two dimensions.

Large faults or faults near the boundaries of the box $\mathcal{B}$ may intersect the edges of $\mathcal{B}$ and are thus not completely contained in the box. This is not accounted for in the calculations below. The probability of a horizontal plane intersecting a fault depends on the vertical thickness T_v. The actual value of T_v is of no interest in the calculations, and for simplicity T_v is assumed to exceed the vertical fault extents.

The variable Y is defined only for faults intersected by the plane, and an indicator variable I_2 of this event is introduced. The discrete stochastic variable I_2 is assigned the value 1 if a fault is intersected by the two dimensional plane, and thus is observed, or 0 otherwise:

$$I_2 = \begin{cases} 1, & \text{if the fault is intersected by the horizontal plane} \\ 0, & \text{else.} \end{cases}$$

The probability of intersection is proportional to the vertical extent of the fault, see Figure 2b, thus larger faults have greater probability of being intersected by the plane. The vertical extent, and hence also the probability of an intersection, are independent of the fault strike. The fault height $h(X, \Theta)$ is given in Expressions (2) and (3), and the probability of the fault intersecting the plane, conditioned on the size $X = x$ and the dip $\Theta = \theta$, is

$$\text{Prob}(I_2 = 1 | X = x, \Theta = \theta) = f_{I_2}(1|x, \theta) = T_v^{-1} h(x, \theta).$$

The probability of the two dimensional plane intersecting an arbitrary fault of unknown size is then found by solving the integral

$$\begin{aligned} f_{I_2}(1) &= \iint f_{I_2}(1|x,\theta) f(x,\theta)\, dx\, d\theta \\ &= T_v^{-1} \iint h(x,\theta) f(x,\theta)\, dx\, d\theta \;=\; T_v^{-1} \mathrm{E}\left[h(X,\Theta)\right]. \end{aligned} \tag{6}$$

The pdf of the size X of a fault intersected by the two dimensional plane is not the same as the original pdf $f(x)$, since the larger faults are over-represented in the two dimensional sample. The new pdf can be found based on the joint pdf of X and Θ for an intersected fault, denoted $f(x, \theta|1)$, which is given by

$$f(x, \theta|1) = \frac{f_{I_2}(1|x,\theta) f(x,\theta)}{f_{I_2}(1)} = \frac{h(x,\theta) f(x,\theta)}{\mathrm{E}\left[h(X,\Theta)\right]}.$$

The fault height is assumed to be on the form $h(x, \theta) = h_1(x) h_2(\theta)$. This is fulfilled by the approximations in Expressions (2) and (3), where the height is given as $h(x, \theta) = c x^a \sin\theta$. Furthermore, X and Θ are assumed to be independent, giving $f(x, \theta) = f(x) f(\theta)$ and $\mathrm{E}\left[h(X,\Theta)\right] = \mathrm{E}\left[h_1(X)\right] \mathrm{E}\left[h_2(\Theta)\right]$. Under these assumptions, the marginal pdf of X of an intersected fault is found to be

$$f(x|1) = \int f(x, \theta|1)\, d\theta = \int \frac{h_1(x) h_2(\theta) f(x) f(\theta)}{\mathrm{E}\left[h_1(X)\right] \mathrm{E}\left[h_2(\Theta)\right]}\, d\theta = \frac{h_1(x) f(x)}{\mathrm{E}\left[h_1(X)\right]}, \tag{7}$$

which is independent of the dip.

The fault size X represents the maximum displacement D or the horizontal extent W. The displacement is at its maximum at the center of the fault, and decreases to zero at the border. The horizontal extent of the ellipsoid is also at its maximum at the center of the fault, and decreases to zero on the top an bottom of the faulted volume. Thus, if the horizontal plane intersects the fault at an arbitrary level away from the center, the observed displacement or horizontal extent at this level are less than D and W respectively. Thus the stochastic variable Y representing the size of the fault observed in the horizontal plane, is equal to or less than X. The variable Y is equal to X only if the plane intersects the fault through its center.

The size Y of a fault observed in an intersecting horizontal plane can be expressed as a function of the size X in three dimensions and the relative vertical distance U from the center of the fault to the horizontal plane, see Expressions (4) and (5). Given that a variable of size X is intersected by the two dimensional plane, the pdf $f(y|x,1)$ is found by a transformation $y = g(u;x)$, where the variable U has a uniform pdf $f(u|x,1)$. The unconditional pdf of Y is found by:

$$\begin{aligned} f(y) &= f(y|1) = \int_y^\infty f(x,y|1)\,dx = \int_y^\infty f(y|x,1)f(x|1)\,dx \\ &= \mathrm{E}\left[h_1(X)\right]^{-1} \int_y^\infty f(y|x,1)h_1(x)f(x)\,dx. \end{aligned} \tag{8}$$

The conditioning on $I_2 = 1$ in the pdf $f(y)$ is omitted, since the variable Y is defined only for intersected faults. Thus $I_2 = 1$ for all faults where the size Y is measured.

The procedure described above can be used to derive the distribution in one dimension, based on the distribution in two dimensions. Some of the faults in the two dimensional plane are then intersected by a line across the plane. In one dimension a fault is only observed as a point on a line, so the fault has no defined width. The size of the fault is therefore represented by the local displacement at the observed point.

Expression (8) gives the forward relationship from the distribution of fault size in three to two dimensions. Previous stereological analysis of ellipsoid or disc shaped faults obtain similar results, and have also focused on the problem of inverting the integral equation (8) in order to derive $f(x)$ based on a known pdf $f(y)$. See for instance Oakeshott and Edwards (1992) and further references therein, Piggott (1997) and Berkowitz and Adler (1998). As pointed out by Oakeshott and Edwards (1992), the inverse problem need not have a unique solution, but may be under- or over-determined. Furthermore, the inverse problem can only be solved analytically in special cases, but must in general be solved numerically. Berkowitz and Adler (1998) present an algorithm for solving the inverse problem for general, non-parametric, $f(x)$. In the current work, the inverse problem is not explicitly solved. The forward calculation in Expression (8) is performed under specified parametric models, and for the distributions $f(y)$ derived in two dimensions at least one solution of the inverse problem exists: the initially specified pdf $f(x)$. In most of the calculations below the distribution class from three dimensions is preserved in two dimensions, and there are reasons to believe that even if there exist other solutions to the inverse problem, the initial pdf $f(x)$ specified in three dimensions is the most plausible one. The results of forward calculations presented below also provide a one-to-one correspondence between the model parameters in two and three dimensions, thus the parameters of the model in three dimensions can be estimated based on observations from two dimensions.

4 Fault Size Distributions

The focus of this work is on fault size distributions. The distributions of interest are the distribution of the displacement D in three, two and one dimensions, and the horizontal extent W in three and two dimensions. First a distribution of D in three dimensions is specified. Through the deterministic relationship between D and W given in Expression (1), the distribution of W in three dimensions can be found based on the distribution of D. By solving the integral (8) for $X = D$ and $X = W$, the distributions of D and W in two and one dimensions can be found. The pdfs of the displacement in three, two and one dimensions are denoted

$$f_3(d), \quad f_2(d) \quad \text{and} \quad f_1(d)$$

respectively, and the pdfs of the horizontal extent, or width, in three and two dimensions are denoted

$$f_3(w) \quad \text{and} \quad f_2(w).$$

Before the fault size distributions are studied, a short discussion on the spatial distribution of faults is given in Section 4.1. In Section 4.2 the cumulative distribution of fault size is discussed in general, relating fault size distributions to common graphical presentations of observed fault sizes. Two possible models for the distributions of fault size in three, two and one dimensions are presented in Sections 4.3 and 4.4, a fractal model and an exponential model respectively.

4.1 Spatial distribution of fault locations

Locations of fault centers $\boldsymbol{\xi}$ are assumed to form a stationary Poisson point process of constant intensity $\lambda(x) = \lambda$, see Stoyan and Stoyan (1994), Stoyan et al. (1995). Conditioned on the number of points, n, the center points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ located within the box $\mathcal{B}$ are independent uniformly distributed in $\mathcal{B}$. The locations of the fault centers are assumed to be independent of the marks of all faults, and the marks of different faults are assumed to be mutually independent. Thus there are no clustering or repulsion of faults, and the fault size is independent of the fault location. When a two dimensional plane intersects $\mathcal{B}$ at an arbitrary depth, the expected number of faults to be intersected by the plane is independent of the position $t \in [0, T_v]$ of the plane. Notice that for a fault with center point outside of the box $\mathcal{B}$, parts of the fault ellipsoid may still intersect $\mathcal{B}$. Thus also faults with center points in a volume surrounding $\mathcal{B}$ have influence on the number of faults in $\mathcal{B}$.

If the distance between the center points of two faults is small, the fault ellipsoids will overlap. In nature, the younger fault may terminate in the older fault, and the resulting

ellipsoid should be truncated. The marked point process does not take any truncation into account. This can lead to a slight sampling bias in two dimensions, since the extents of faults that should be truncated are slightly too large in the specified model.

4.2 Cumulative fault size distribution

Consider the stochastic variable X representing the size of a fault, with pdf $f(x)$. The function $S(x)$ is defined in the following way:

$$S(x) = \text{Prob}(X \geq x) = \int_x^\infty f(u)\, du, \tag{9}$$

hence $S(x) = 1 - F(x)$ where $F(x)$ is the cumulative distribution function of X. If the number of observed faults is n, with a corresponding set of observations x_i, $i = 1, 2, \ldots, n$, an empirical estimate of the function $S(x)$ is given by

$$\hat{S}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}[x_i \geq x], \tag{10}$$

where $\mathbf{1}[\cdot]$ is the indicator function. The estimate of $S(x)$ is thus given as the relative number of faults with size x_i equal to or larger than x. Observations of fault sizes are often presented on a bilogarithmic plot, hereafter referred to as a log-log plot, giving empirical plots of $\log \hat{S}(x)$ versus $\log x$. An example of a log-log plot of observed displacements is shown in Figure 5. Instead of plotting the relative number of faults on log-log scale, the absolute number is often plotted, thus giving an estimate $n\hat{S}(x)$. In some cases the plots are normalized to give the number of faults per km^2 in two dimensions or per km in one dimension.

4.3 Fractal distributions

The term fractal was introduced by Mandelbrot (1983), and an introduction to fractals is also given by Feder (1988). The distribution of the stochastic variable X, representing the maximum displacement or width of a fault, is called fractal if it follows a scaling relationship:

$$f(\lambda x) = \lambda^H f(x),$$

where H is the scaling exponent, see Feder (1988). An example of a scaling distribution is the negative power law distribution

$$S(x) = Cx^{-\beta}, \qquad \beta > 0, \tag{11}$$

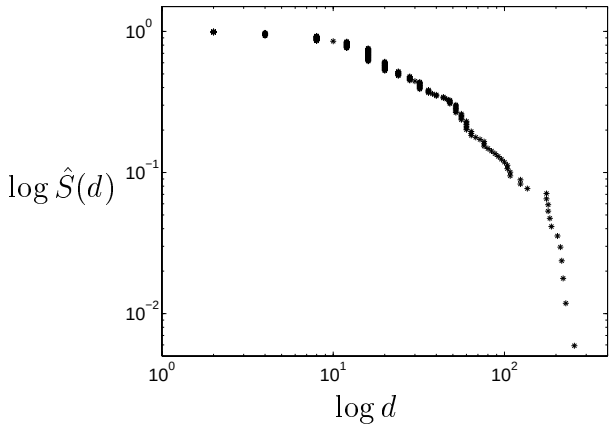


Figure 5: Log-log plot of relative number $\hat{S}(d)$ of faults with observed displacements larger than d, plotted at observations points d_i, $i = 1, \ldots, n$.

where C is a positive constant and S(x) is defined in Expression (9). The parameter β is called the fractal dimension or the power law exponent. Negative power law distributions are frequently used in geology to model distributions of fault displacements and widths, see for example Childs et al. (1990), Heffer and Bevan (1990), Walsh et al. (1991), Marrett and Allmendinger (1992), Gauthier and Lake (1993) and Odling (1997). See also JSG (1996, Vol. 18).

The function $S(x)$ is often estimated as the relative number of observations larger than x, $\hat{S}(x)$, see Expression (10). If X follows the negative power law (11) and is observed exactly, a plot of $\hat{S}(x)$ versus x should produce a straight line of slope $-\beta$ on a log-log plot.

The relative number $\hat{S}(x)$ should be bounded above by 1. According to the negative power law (11) however, as $x \to 0$, the function $S(x)$ increases unlimited: $S(x) \to \infty$. Allowing $x = 0$, no constant C can be chosen to satisfy the upper bound condition. Thus it is not possible to define a negative power law distribution for all $X \geq 0$, but a lower limit $x_0 > 0$ for X must be chosen. Above this lower bound the distribution of X is fractal, while below the limit the distribution is left unspecified. Thus the need of a lower bound results in a model which is not able to reflect the true distribution on all scales. The choice of x_0 can be related to the geological properties of the rock, and a possible lower limit is the grain size of the minerals. In practice, an artificial lower bound is often used, for example the limit of resolution for a set of observations.

The pdf of a stochastic variable X following the negative power law distribution (11) is of the form

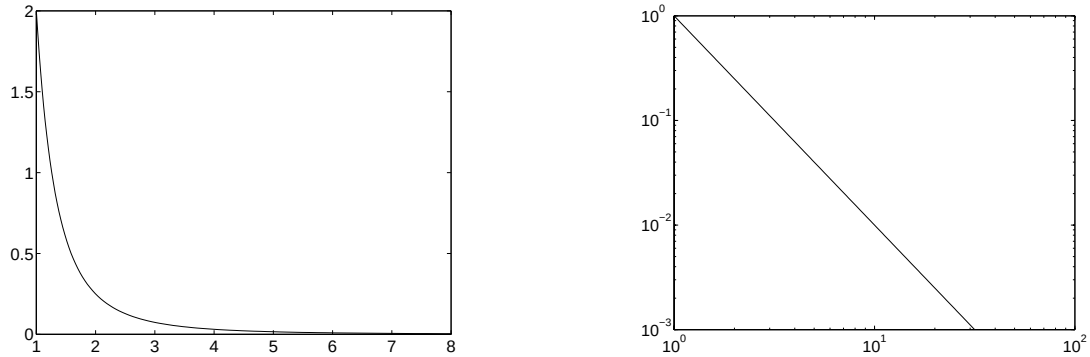
$$f(x; x_0, \beta) = \frac{\beta x_0^\beta}{x^{\beta+1}}, \qquad x \geq x_0,\ x_0 > 0,\ \beta > 0, \tag{12}$$

which is the pdf of the Pareto distribution (Johnson et al., 1994). The lower limit $x_0 > 0$,

discussed above, ensures that the integral of the pdf is finite. The pdf is plotted in Figure 6a, while Figure 6b shows a log-log plot of the function $S(x)$, given as

$$S(x) = \left(\frac{x_0}{x}\right)^{\beta}, \qquad x \geq x_0. \tag{13}$$

Suppose a lower limit x_{03} is used in a negative power law distribution of the size of faults in three dimensions. When the corresponding distribution of fault size in two dimensions is studied, only faults of size above a lower limit $x_{02} \geq x_{03}$ in two dimensions can be considered. If a fault observed in two dimensions has a size smaller than x_{03}, it is uncertain if the size of the fault in three dimensions is smaller or larger than x_{03}, and thus the distribution of the fault size in three dimensions is unknown. By the same reasoning a lower limit x_{01} in one dimension must be defined, satisfying $x_{01} \geq x_{02}$.



(a) (b)

Figure 6: The Pareto distribution with parameters $\alpha = 1.0$ and $\beta = 2.0$. Figure (a) shows the pdf $f(x)$, Expression (12), and figure (b) the function $S(x)$, Expression (13), on log-log scale.

4.3.1 Displacement

The pdf of the displacement in three dimensions is given by

$$f_3(d) = \frac{\beta_3 d_{03}^{\beta_3}}{d^{\beta_3+1}}, \qquad d \geq d_{03}, \tag{14}$$

which is of the form given in Expression (12), where the lower bound and power law exponent are denoted d_{03} and β_3 respectively. Let D' denote the displacement obtained in two dimensions. The procedure described in Section 3 is used, with the approximation of the height of the fault ellipsoid given in Expression (3). Then, for $D' \geq d_{03}$ the unconditional pdf of D' is found to be

$$f_{D'}(d') = \frac{\beta_3 - 1/\tau}{\gamma(\beta_3 - 1/\tau) + 1} \cdot \frac{d_{03}^{\beta_3 - 1/\tau}}{(d')^{\beta_3 - 1/\tau + 1}}, \tag{15}$$

where γ is the parameter involved in the displacement function (5). The calculations are given in Appendix B. As pointed out above, a lower limit $d_{02} \geq d_{03}$ should be used for the displacement D' obtained in two dimensions. A distribution with this lower bound is derived by conditioning on $D' \geq d_{02}$. The conditional pdf can be found from Expression (15), without knowing the pdf for $D' < d_{03}$. The two parameters β_3 and τ only appear in the combination $\beta_3 - 1/\tau$ in Expression (15) and in the corresponding conditional pdf. Thus the pdf in two dimensions can be expressed as a two-parameter distributions with parameters $\beta_2 = \beta_3 - 1/\tau$ and d_{02}, and is found to be

$$f_2(d) = f_{D'}(d|d \geq d_{02}) = \frac{\beta_2 d_{02}^{\beta_2}}{d^{\beta_2+1}}, \qquad d \geq d_{02}. \tag{16}$$

This pdf is of the form given in Expression (12), and the distribution of the displacement is fractal also in two dimensions.

The distribution of maximum displacement in one dimension can only be found if the relationship between D and W in two dimensions is known. This relationship is discussed below, after obtaining the distribution of fault width in two dimensions. Under the assumption that the relationship in three dimensions, Expression (1), also holds in two dimensions, the distribution in one dimension is found by the same procedure as above. Assuming independence between fault strike and displacement, the pdf of D in one dimension is found as

$$f_1(d) = \frac{\beta_1 d_{01}^{\beta_1}}{d^{\beta_1+1}}, \qquad d \geq d_{01}, \tag{17}$$

where $\beta_1 = \beta_3 - 2/\tau$ and $d_{01} \geq d_{02}$. A negative power law is obtained also in one dimension. This result is only valid if $\beta_3 > 2/\tau$. If $\beta_3 \leq 2/\tau$, non-tractable expressions arise in the calculations.

4.3.2 Width

Expression (1) gives the relationship $W = (Dc_1)^{1/\tau}$ between fault width and displacement. Thus the pdf of W in three dimensions can be found from the pdf (14), by transformation of D, giving the fractal distribution

$$f_3(w) = \frac{\mu_3 w_{03}^{\mu_3}}{w^{\mu_3+1}}, \qquad w \geq w_{03}, \tag{18}$$

where $w_{03} = (d_{03}c_1)^{1/\tau}$ is the lower limit of W and $\mu_3 = \tau\beta_3$. Let W' denote the width obtained in two dimensions. The procedure described in Section 3, with the approximation (2) of the fault heights, gives the following pdf of W' for $W' \geq w_{03}$:

$$f_{W'}(w') = \frac{\Gamma((\mu_3+1)/2)}{\Gamma((\mu_3+2)/2)} \cdot \frac{\sqrt{\pi}}{2} \cdot \frac{(\mu_3 - 1) w_{03}^{\mu_3 - 1}}{(w')^{\mu_3}}. \tag{19}$$

The calculations are given in Appendix B. As for the displacement, a lower limit w_{02} in two dimensions should be defined for the width, satisfying $w_{02} \geq w_{03}$. The pdf of the width in two dimensions, now denoted W, is found from the pdf (19), conditioning on $W \geq w_{02}$:

$$f_2(w) = f_{W'}(w|w \geq w_{03}) = \frac{(\mu_3 - 1)w_{02}^{\mu_3 - 1}}{w^{\mu_3}} = \frac{\mu_2 w_{02}^{\mu_2}}{w^{\mu_2 + 1}}, \qquad w \geq w_{02}, \tag{20}$$

where $\mu_2 = \mu_3 - 1 = \tau\beta_3 - 1$. This pdf is of the form given in Expression (12), and thus also the width distribution is fractal in two dimensions.

As pointed out in Section 4.3.1, the relationship between W and D in two dimensions is needed to obtain $f_1(d)$ from $f_2(d)$. In three dimensions the relationship $W = (Dc_1)^{1/\tau}$ applies. The relationship between width W' and maximum displacement D' of the ellipse observed in an intersection plane is obtained by combining Expressions (4) and (5). If the plane intersects the fault ellipsoid at relative distance $Uh(d,\theta)/2$ from the center point, where U is uniformly distributed on $[-1,1]$, the relationship is given as

$$W' = \sqrt{1 - U^2}(1 - |U|)^{-\gamma/\tau}(D'c_1)^{1/\tau}.$$

If the mean value with respect to U is calculated, the relationship $W' = (D'c_1')^{1/\tau}$ is obtained, assuming the expression is integrable. This relationship is consistent with the transformation

$$f_2(d) = f_2(w(d)) \cdot |w'(d)|,$$

using $f_2(d)$ from Expression (16) and $f_2(w)$ from Expression (20), and is used above when the pdf of the displacement in one dimension is found.

4.4 Exponential distributions

The fractal distribution is widely used to model the size distribution of faults. In some cases however, fault sizes are observed to be better described by an exponential distribution. For example, exponential distribution of fault size are observed on mid-ocean ridges, see Cowie et al. (1994), and transformations from fractal distribution at early stages of deformation into exponential distributions are observed in analog experiments of fault growth, see Spyropoulos et al. (1999) and Ackermann et al. (1999). The pdf of the exponential distribution is

$$f(x;\lambda) = \lambda \exp(-\lambda x), \qquad x \geq 0,\ \lambda > 0, \tag{21}$$

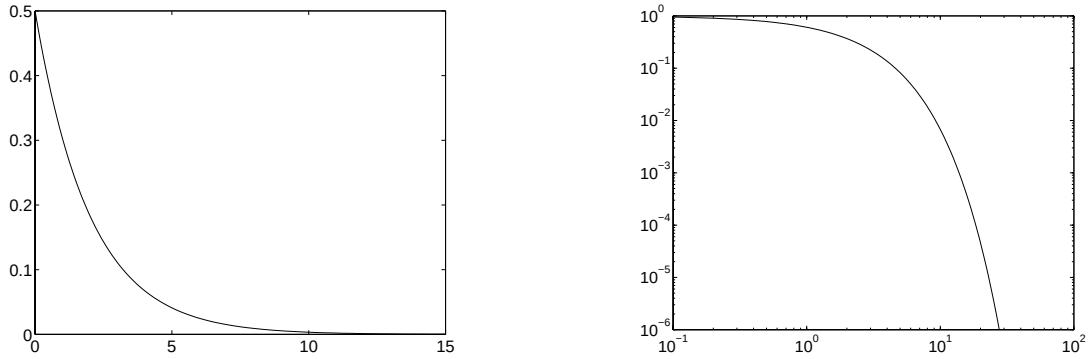
and Figure 7a shows a plot of the function. For this distribution there is no need to specify a lower limit for x, as was the case for the fractal distribution. The function $S(x)$, defined in Expression (9), is

$$S(x) = \exp(-\lambda x), \qquad x \geq 0, \tag{22}$$

and is plotted on log-log scale in Figure 7b. For the exponential distribution, the integrals which have to be solved to derive the size distributions in lower dimensions, see Section 3, are not analytical tractable for general τ and γ. Some expressions for general parameter values are obtained below, but final results are obtained only under the assumption

$$\tau = \gamma = 1. \tag{23}$$

This value of τ corresponds to a linear relationship between maximum displacement and fault width, Expression (1), while the value of γ implies a liner displacement function, see Expression (5). Under this requirement, the integrals involved in the calculations below are analytically tractable.



(a) (b)

Figure 7: The exponential distribution with parameter $\lambda = 0.5$. Figure (a) shows the pdf $f(x)$ in Expression (21) and figure (b) the function $S(x)$ in Expression (22) on log-log scale.

4.4.1 Displacement

The pdf of the displacement in three dimensions is given by

$$f_3(d) = \lambda \exp(-\lambda d), \qquad d \geq 0. \tag{24}$$

Using the procedure described in Section 3, and the approximation (3) of the height of the fault ellipsoids, the pdf of D in two dimensions is found to be

$$f_2(d) = \gamma^{-1} \lambda^{1/\gamma} d^{1/\gamma - 1} \frac{\Gamma(1/\tau - 1/\gamma + 1, \lambda d)}{\Gamma(1 + 1/\tau)}, \qquad d \geq 0, \tag{25}$$

where the parameters τ and γ must satisfy $\gamma > \tau/(1+\tau)$, see Appendix B. The incomplete gamma function $\Gamma(\cdot, \cdot)$ is described in Appendix C. Under the assumption in Expression (23), the pdf simplifies to

$$f_2(d) = \lambda \exp(-\lambda d), \qquad d \geq 0, \tag{26}$$

which is the pdf of the exponential distribution, see Expression (21). Thus under the specified assumptions, if the distribution of D in three dimensions is exponential, then the distribution is exponential also in two dimensions. The parameter is λ both in three and two dimensions.

As in Section 4.3.1, the relationship between W and D in two dimensions is required to be able to derive $f_1(d)$. This relationship is discussed below. If the relationship $W \propto D$ is assumed to be valid also for displacement and width in two dimensions, the displacement has the same exponential distribution in one dimension as in three and two:

$$f_1(d) = \lambda \exp(-\lambda d), \qquad d \geq 0. \tag{27}$$

4.4.2 Width

The pdf of the width in three dimensions is found from the pdf of D in Expression (24), using the transformation $W = (Dc_1)^{1/\tau}$:

$$f_3(w) = \rho\tau w^{\tau-1} \exp(-\rho w^\tau), \qquad w \geq 0, \tag{28}$$

where the parameter ρ is given by $\rho = \lambda c_1^{-1}$. This is the pdf of a Weibull distribution. Let W' denote the width obtained in two dimensions. Using the procedure in Section 3 and the approximation (2) of the fault height, the pdf of W' in two dimensions is found to be

$$f_{W'}(w') = \frac{\tau\rho^{1+1/\tau}}{\Gamma(1+1/\tau)} w' \int_{w'}^{\infty} \frac{w^{\tau-1}}{\sqrt{w^2 - (w')^2}} \exp(-\rho w^\tau)\, dw, \tag{29}$$

see Appendix B. The integral does not have a general analytical solution for arbitrary τ. Returning to the assumption in Expression (23), the pdf of W in three dimensions, Expression (28), becomes $f_3(w) = \rho \exp(-\rho w)$, the pdf on an exponential distribution. The pdf of the width in two dimensions, now denoted W, is found from Expression (29):

$$f_2(w) = \rho^2 w K_0(\rho w), \qquad w \geq 0, \tag{30}$$

see Appendix B. The mean and variance of W are $\mathrm{E}[W] = \pi/2\rho$ and $\mathrm{Var}[W] = (16 - \pi^2)/4\rho^2$. The function $K_0(\cdot)$ is the modified Bessel function of order 0, see Appendix C, and the pdf $f_2(w)$ is shown in Figure 8a. Unlike the pdfs of the Pareto distribution and the exponential distribution, this function is not monotonely decreasing. The function $S(w)$ defined in Expression (9) is in this case given by

$$S(w) = \rho w K_1(\rho w), \qquad w \geq 0, \tag{31}$$

where $K_1(\cdot)$ is the modified Bessel function of order 1, see Appendix C. The function $S(w)$ is plotted on log-log scale in Figure 8b.

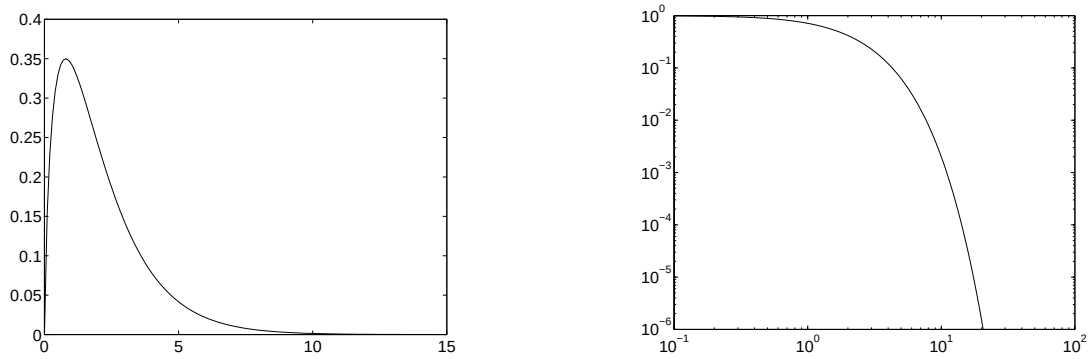
The relationship between W and D in two dimensions was needed in the derivation of $f_1(d)$ from $f_2(d)$, and the pdf in Expression (27) was derived under the assumption $W \propto D$. This relationship is however not consistent with the transformation

$$f_2(d) = f_2(w(d)) \cdot |w'(d)|, \tag{32}$$

using $f_2(d)$ from Expression (26) and $f_2(w)$ from Expression (30). It can be shown that this transformation is obtained for the relationship

$$D = -\lambda^{-1} \ln(\rho W K_1(\rho W)).$$

This relationship is slightly convex, but a relatively good approach to an asymptotic linearity is soon established as W increases.



(a) (b)

Figure 8: (a) The pdf $f_2(w)$ given in Expression (30) with parameter $\rho = 0.75$. (b) The function $S(w)$ in Expression (31) on log-log scale.

4.4.3 Exponential distributions with lower bounds

The pdf of a fractal distribution only exists if a lower limit of fault size is specified. This lower bound is not necessary for the distributions derived from an exponential distribution of displacement in three dimensions. However, to compare these distributions with the fractal distributions, the same lower bounds are introduced. Under the assumption in Expression (23), equal exponential distributions are obtained for the displacement in three, two and one dimensions, see Expressions (24), (26) and (27). Introducing a lower bound d_0, where $d_0 = d_{03}$ in three dimensions, d_{02} in two and d_{01} in one, the pdf becomes

$$f(d) = \lambda \exp(-\lambda(d - d_0)), \qquad d \geq d_0. \tag{33}$$

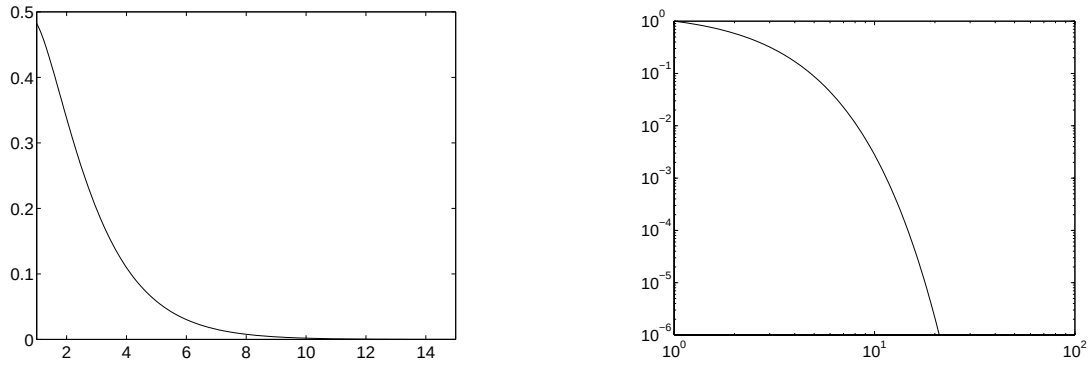
The pdfs of the width distributions in Expressions (28) and (30), with the same lower bounds as for the fractal distributions, are found to be

$$\begin{aligned} f_3(w) &= \rho \exp(-\rho(w - w_{03})), \qquad w \geq w_{03}, \\ f_2(w) &= \rho \cdot \frac{wK_0(\rho w)}{w_{02}K_1(\rho w_{02})}, \qquad w \geq w_{02}. \end{aligned} \tag{34}$$

The pdf (34) is shown in Figure 9a. For reasonably large values of the lower bound w_{02}, the truncated pdf is monotonely decreasing. When lower bounds are imposed on the width, the function $S(w)$ corresponding to the pdf (34) is

$$S(w) = \frac{wK_1(\rho w)}{w_{02}K_1(\rho w_{02})}, \tag{35}$$

and is illustrated on log-log scale in Figure 9b.



(a) (b)

Figure 9: (a) The pdf $f_2(w)$ given in Expression (34) width parameter $\rho = 0.75$ and lower bound $w_{02} = 1.0$. (b) The corresponding function $S(w)$ in Expression (35) on log-log scale.

4.5 Number of faults

In addition to the fault size distributions, the distribution of the number of faults is also of interest. The fault positions are assumed to follow a stationary Poisson point process, see Section 4.1, thus the distribution of the number of faults, K, is a Poisson distribution:

$$f(k;\omega) = \frac{\omega^k}{k!}\, e^{-\omega}, \qquad \omega > 0,\ k = 0, 1, 2, \ldots, \tag{36}$$

where ω is the intensity of the Poisson process. The number of faults in both three, two and one dimensions will follow a Poisson distribution, but with different intensity ω in different dimensions.

For seismic data, prior information about fault intensity can be obtained from wells, providing information about the number of faults along a one dimensional line. Thus if data is obtained in two dimensions, the fault intensity obtained from the well should be converted to an approximate number of faults in the area of observations. Denote by K_2 the number of faults in the two dimensional observation area and by K_1 the number intersected by a one dimensional line. The expected number of faults along a line through the area is then given as $\mathrm{E}\left[K_1|K_2\right] = f_{I_1}(1)K_2$, where $f_{I_1}(1)$ is the probability of a fault in the area being intersected by the line. If k_1 is obtained from well information, a rough estimate of K_2 given as $\hat{k}_2 = k_1/f_{I_1}(1)$. The probability $f_{I_1}(1)$ is calculated in a similar way as the probability f_{I_2} of a fault being intersected by a two dimensional plane, see Expression (6), and is proportional to the mean value of the fault width in two dimensions. Thus based on one dimensional well observations, the prior mean number of faults in two dimensions, $\mathrm{E}\left[K_2\right] = \omega$, should be chosen as

$$\omega = \text{const} \times \frac{k_1}{c_1 \mathrm{E}_2\left[D\right]}, \tag{37}$$

under the assumption $\tau = 1$ in Expression (23), where $\mathrm{E}_2\left[D\right]$ is the expected maximum displacement in two dimensions. The constant term is related to the size of the observation area. Since $\mathrm{E}_2\left[D\right]$ differs under the fractal and exponential model, the mean number of faults ω should be chosen differently for each model, if based on the same fault intensity in one dimension. Note that the mean value of D in two dimensions in the case of a Pareto distribution is only finite for $\beta_2 > 1$, see Expression (16), thus Expression (37) only holds in this case. If $\beta_2 \leq 1$, the number of faults in two dimensions can be approximated by only considering faults of displacement below some specified size, $D \leq d_{\max}$, in which case the mean value is finite under both models.

4.6 Prior distribution

The distributions presented in Sections 4.3–4.5 can serve as prior distributions for a fault population, when a Bayesian framework is used to make inference based on observations of faults. Fault sizes in an observation area are now denoted $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$, where Y_j represents either the maximum displacement or the width of a fault. The prior pdf of the number of faults in the observation area, K, and their sizes $\boldsymbol{Y}$ can be expressed as

$$f(\boldsymbol{y}, k, \theta) = f(\boldsymbol{y}|k, \theta) f(k; \omega) f(\theta), \tag{38}$$

where Θ is a model parameter involved in the pdf of $\boldsymbol{Y}$, while K has the prior pdf defined in Expression (36) and is independent of Θ. Assuming Y_j, $j = 1, \ldots, K$, are iid variables, the pdf $f(\boldsymbol{y}|k, \theta)$ is given as

$$f(\boldsymbol{y}|k, \theta) = \prod_{j=1}^{k} f(y_j|\theta) \tag{39}$$

where $f(y_j|\theta)$ is the marginal pdf of the fault size Y_j. If the observation area is two dimensional, the pdfs given in Expression (16), (26) or (33) can be used for the maximum displacement. For fault widths observed in two dimensions the pdfs in Expression (20), (30) or (34) can be used as prior pdfs. Some proper prior $f(\theta)$ is chosen for the model parameter. The prior pdf of $(\boldsymbol{Y}, K, \theta)$ is thus

$$f(\boldsymbol{y}, k, \theta) = \frac{\omega^k}{k!}\, e^{-\omega} f(\theta) \prod_{j=1}^{k} f(y_j|\theta). \tag{40}$$

5 Observations of Faults

Faults can be directly observed in outcrops or in core samples from wells, or indirectly observed from seismic data. In either case there is a sampling error connected to the observations, see for example Heffer and Bevan (1990) and Pickering et al. (1995). Different types of errors are of interest in this work: measurement errors, censoring and missing observations. A model for observation errors is defined through a likelihood function below. Qualitatively the model is based on the discussions of error types given by Heffer and Bevan (1990) and Pickering et al. (1995), although the cited papers do not use the same parametric models.

For seismic data, measurement errors mainly consist of a downward bias, where both fault widths and maximum displacements tend to be measured too small, due to the limit of resolution. Furthermore, the resolution of seismic data also results in a number of faults not being observed, even at scales above the limit of resolution, hence there are often missing observations. In geophysical literature, the term truncation is sometimes used to denote both the downward bias in measured fault width and the missing observations at some scales. Censoring of fault size observations is a result of the bounded observation area. For faults extending beyond the observation area, the observed sizes are censored and only give the size observed within the area.

5.1 Likelihood function

Consider the set of faults within an observation area, with sizes $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$. Not all faults are observed, and the stochastic vector $\boldsymbol{S} = (S_1, \ldots, S_K)$ of indicator variables $S_j \in \{0, 1\}$ describes which faults that are observed. If fault j is observed, then $S_j = 1$, while $S_j = 0$ otherwise. The number of observed faults is thus given as $N = \sum_{j=1}^{K} S_j \leq K$. Furthermore, due to measurement errors, the observed values, denoted $\boldsymbol{Z}$, are not exact. The likelihood function of the observations accounts for the observation errors described above, and include information about the likelihood of observing the N faults defined by the indicator vector $\boldsymbol{S}$, with observed sizes $\boldsymbol{Z}$, when the actual number of faults is K and

their sizes are $\boldsymbol{Y}$. The probability of observing a fault of size Y_j is given as $p(Y_j)$, depending on the fault size, but independent of all other faults. Thus $\mathrm{P}\{S_j = 1|Y_j = y_j\} = p(y_j)$. If a fault is observed, i.e., $S_j = 1$, the observed value Z_j is described by the likelihood function $h(z_j|y_j)$, depending only on Y_j since Z_j is a measurement of this specific fault. The likelihood function of the observations $\boldsymbol{X} = (\boldsymbol{Z}, \boldsymbol{S})$ can thus be expressed as

$$h(\boldsymbol{x}|\boldsymbol{y}, k) = h(\boldsymbol{s}|\boldsymbol{y}, k)h(\boldsymbol{z}|\boldsymbol{s}, \boldsymbol{y}, k) = \prod_{j=1}^{k} p(y_j)^{s_j}(1 - p(y_j))^{1-s_j} \prod_{j;s_j=1} h(z_j|y_j). \tag{41}$$

5.1.1 Observed number of faults

For a seismic dataset $\boldsymbol{x} = (\boldsymbol{z}, \boldsymbol{s})$ there are typically missing observations, and only a proportion of the true number of faults is observed. The probability of observing exactly the n faults observed, conditioned on K and $\boldsymbol{Y}$, is given as

$$h(\boldsymbol{s}|\boldsymbol{y}, k) = \prod_{j=1}^{k} p(y_j)^{s_j}(1 - p(y_j))^{1-s_j}$$

and is specified through the function $p(y)$. It is natural to assume that $p(y)$ is a monotonely increasing function, see Clark et al. (1999), since missing observations due to limited resolution of the measurement equipment is most dominant for small scale faults, while the presence of large scale faults in the observation area is more frequently detected. In this work $p(y)$ is assumed to be piecewise linear, given as

$$p(y) = \begin{cases} p_1, & y \leq \eta_1 \\ (\eta_2 - \eta_1)^{-1}((p_2 - p_1)y + p_1\eta_2 - p_2\eta_1), & \eta_1 \leq y \leq \eta_2 \\ p_2, & y > \eta_2, \end{cases} \tag{42}$$

and illustrated in Figure 10. The function gives a constant small probability p_1 of observing faults of size smaller than $y = \eta_1$, and a constant high probability p_2 of observing faults of size larger than $y = \eta_2$. Between $y = \eta_1$ and $y = \eta_2$ the probability of observation is linearly increasing. The parameter η_1 is related to the limit of resolution, and p_1 can be interpreted as the probability of observing a fault of size at the limit of resolution. The parameter η_2 can be interpreted as a value for which most faults of size larger than this value will be observed. In the special case $p_1 = p_2 = 1$ all faults have probability 1 of being observed, thus $S_j = 1$, $j = 1, \ldots, K$, and the observed number of faults equals the true number, $n = K$.

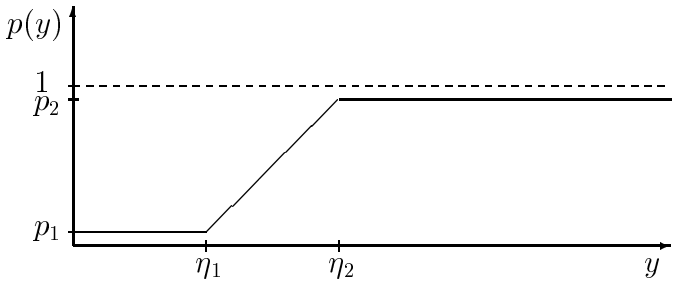


Figure 10: The function $p(y)$ in Expression (42).

5.1.2 Observed fault sizes

The likelihood function of observed fault sizes $\boldsymbol{Z}$,

$$h(\boldsymbol{z}|\boldsymbol{s},\boldsymbol{y},k) = \prod_{j;s_j=1} h(z_j|y_j),$$

is used to model the measurement errors, which typically leads to a downward bias. For an observed fault, the measurement errors is given as the difference $Y_j - Z_j$. For seismic observations, the observed value tends to give a too small measurement of Y_j, due to the resolution of the seismic data, thus $Y_j - Z_j \geq 0$. For simplicity, it is assumed to be a lower and an upper limit of possible measurement error of this type, thus $\delta_1 \leq Y_j - Z_j \leq \delta_2$, and a uniform distribution of the error within this interval is chosen.

The dimension of the observation area of faults enforces an upper bound η on observable fault size Z. If fault widths are observed, η is directly related to the size of the observation area. There is however no strict upper bound of observable maximum displacement inside a bounded area, hence in this case η can be chosen as $\eta = \infty$. If the true fault size is only slightly larger than η, the downward bias in the observations gives an observation $Z \leq \eta$ anyway. As Y increases, the observed value is censored and remains below η. The observation Z now only measures the size of the fault within the observation area, and some bias is still allowed, with the same lower and upper limits as before.

If Y exceeds the upper bound η, censoring is guaranteed. However, also smaller faults, $Y < \eta$, can in practice be censored, if one or both fault tips cross the boundary of the observation area. This kind of censoring is not accounted for in the model.

The uniform likelihood function $h(z|y)$ is illustrated in Figure 11, and can be expressed as

$$h(z|y) = \frac{1 + (\delta_2 - y)\delta_z(0)}{\delta_2 - \delta_1}, \qquad \delta_1 \leq \min\{y, \eta + \delta_1\} - z \leq \delta_2 \text{ and } z \geq 0, \tag{43}$$

where the term $(\delta_2 - y)\delta_z(0)$; $\delta_z(\cdot)$ being the Dirac delta function, is included to account for boundary effects for small fault sizes. The choice of edge correction has only minor influence on the model.

The special case $\delta_1 = \delta_2 = 0$ implies no bias in the observations, and the fault size is observed exactly for all faults with size below the upper bound η. In this case the likelihood function is replaced by $h(z|y) = \delta_z(\min\{y, \eta\})$. If in addition the upper bound is $\eta = \infty$, all fault sizes are observed exactly.

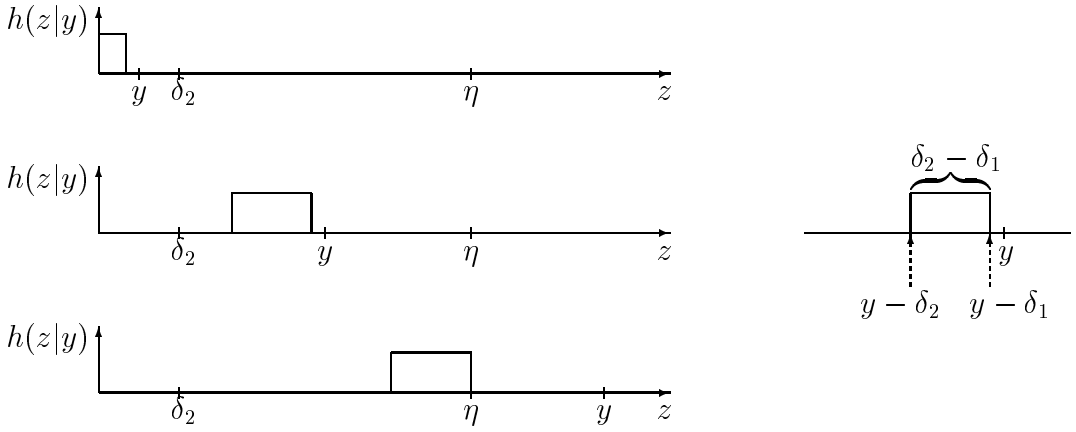


Figure 11: The likelihood function $h(z|y)$, Expression (43), for different values of Y. The left hand figures illustrate the uniform distribution of Z for small, intermediate and large values of Y. The right hand figure shows the interval boundaries of the distribution.

5.2 Posterior distribution

The focus of this work is on the posterior distribution of fault sizes, conditioned to observations, and inference about the underlying fault population is based on this distribution. In a Bayesian setting, the prior pdf $f(\boldsymbol{y}, k, \theta) = f(\boldsymbol{y}, k|\theta)f(\theta)$ of $(\boldsymbol{Y}, K, \Theta)$, see Section 4.6, includes general geological knowledge about the distribution of fault sizes and the number of faults in the observation area. The likelihood function $h(\boldsymbol{x}|\boldsymbol{y}, k)$ of the observations $\boldsymbol{X} = (\boldsymbol{Z}, \boldsymbol{S})$ models the data collection, see Section 5.1. The observations are assumed to be conditionally independent of the parameter Θ, given $(\boldsymbol{Y}, K)$. Combining the prior pdf (40) of the true fault population $(\boldsymbol{Y}, K, \theta)$ and the likelihood function (41) of the

observations $\boldsymbol{X} = (\boldsymbol{Z}, \boldsymbol{S})$, the posterior pdf is obtained:

$$\begin{aligned} f(\boldsymbol{y}, k, \theta|\boldsymbol{x}) &= \text{const} \times \frac{\omega^k}{k!} e^{-\omega} f(\theta) \prod_{j=1}^{k} f(y_j|\theta) \\ &\quad \times \prod_{j=1}^{k} p(y_j)^{s_j} (1 - p(y_j))^{1-s_j} \prod_{j;s_j=1} h(z_j|y_j). \end{aligned} \tag{44}$$

6 Model Choice

The relationship between the observations $\boldsymbol{X}$ and the true variables $(\boldsymbol{Y}, K)$ is described by the likelihood function in Expression (41). Suppose now that the pdf $f(\boldsymbol{y}, k|\theta)$, Section 4.6, is unknown, but two possible models are suggested. The stochastic variable $M \in \Omega_m = \{m_1, m_2\} = \{1, 2\}$ indicates which model is the correct one, model 1 or 2 respectively. Under model m, the pdf of $(\boldsymbol{Y}, K)$ is denoted $f(\boldsymbol{y}, k|m, \theta_m) = f(k|m) f(\boldsymbol{y}|m, k, \theta_m)$, $m = 1, 2$, where different model parameters θ_m with prior pdfs $f(\theta_m|m)$ describe the two distributions. Based on the observations $\boldsymbol{X}$ it is of interest to determine which of the two pdfs of $(\boldsymbol{Y}, K)$ is most likely. This can be expressed as a model choice problem, and Bayes factor can be used as a criterion for choosing the model best fitted to the observations. See Kass and Raftery (1995) for an overview. The two competing models are assigned prior probabilities $f(m_1)$ and $f(m_2)$. Bayes factor $B_{12}(\boldsymbol{x})$ of the observations $\boldsymbol{X}$ is defined as

$$B_{12}(\boldsymbol{x}) = \frac{f(m_1|\boldsymbol{x})}{f(m_2|\boldsymbol{x})} \cdot \frac{f(m_2)}{f(m_1)} = \frac{f(\boldsymbol{x}|m_1)}{f(\boldsymbol{x}|m_2)}, \tag{45}$$

where $f(m|\boldsymbol{x})$ is the posterior probability of model m. If the value of Bayes factor exceeds one, $B_{12}(\boldsymbol{x}) > 1$, model 1 is favored, while for a value smaller than one model 2 is most likely. If the prior probabilities of model 1 and 2 are equal, Bayes factor coincides with the posterior odds ratio. The marginal pdf of $\boldsymbol{X}$ under model m is given as

$$f(\boldsymbol{x}|m) = \sum_{k=0}^{\infty} \iint h(\boldsymbol{x}|\boldsymbol{y}, k) f(\boldsymbol{y}, k|m, \theta_m) f(\theta_m|m) \, d\theta_m \, d\boldsymbol{y}. \tag{46}$$

In some cases Bayes factor can be found analytically, but in general the value must be estimated by the use of numerical integration techniques. Evans and Swartz (1995) and Kass and Raftery (1995) discuss different estimation techniques. In this work, Bayes factor is estimated by the use of McMC sampling, where the model M is included as a stochastic variable, see Carlin and Chib (1995). An introduction to McMC theory is given, for instance, by Besag et al. (1995) and Han and Carlin (2000) give a review of McMC methods for estimating Bayes factor.

Borgos (2000) describes an McMC sampling algorithm constructed to estimate Bayes factor based on non-exact observations. The McMC algorithm introduces auxiliary variables K_1

and K_2, representing the number of faults K under the two competing models. The posterior distribution of $(M, K_1, K_2, \Theta_1, \Theta_2)$ given $\boldsymbol{X}$ can be found, integrating over $\boldsymbol{Y}$. The McMC algorithm then generate samples from this posterior distribution, and the variable $\boldsymbol{Y}$ need not be included in the McMC simulations. Based on realizations $(m^j, k_1^j, k_2^j, \theta_1^j, \theta_2^j)$, $j = 1, 2, \ldots, r$, generated by the McMC algorithm, Bayes factor can be estimated as

$$\hat{B}_{12}(\boldsymbol{x}) = \frac{\sum_{j=1}^{r} \mathbf{1}[m^j = m_1]}{\sum_{j=1}^{r} \mathbf{1}[m^j = m_2]} \cdot \frac{f(m_2)}{f(m_1)}, \tag{47}$$

where $\mathbf{1}[\cdot]$ is the indicator function. Estimation of the model parameter Θ_m, the number of faults K and fault sizes $\boldsymbol{Y}$ under the favored model is discussed in Section 6.2.

6.1 Competing models

Two alternative fault size distributions are discussed in Section 4, a fractal and an exponential model. As illustrated in Figure 6b, for the fractal model a log-log plot of the function $S(x)$, Expression (9), gives a straight line. However, the log-log plot of the estimate $\hat{S}(x)$, Expression (10), based on observed fault sizes often shows a concave shaped curve. The concave shape can in some cases be explained by the typical sampling errors, especially missing observations can result in this shape of the curve. However, the concave curve can also be a result of a different underlying distribution, for example an exponential model. As shown in Figure 7b, an exponential distribution gives a concave curve on a log-log plot of $S(x)$. The aim of this work is to use a statistical framework to investigate which model is the most likely one, based on a set of fault observations.

In the calculations below, fault size Y is assumed to be measured as maximum displacement D. True fault sizes $\boldsymbol{Y}$ are assumed to be iid variables, see Expression (39), while the number of faults K follows a Poisson distribution, Expression (36). The fractal model with the pdf in Expression (16) as marginal distribution of Y_j and model parameter $\theta_1 = \beta_2$ is used as model 1. The assumption in Expression (23) is assumed to hold, and the truncated exponential model with the pdf in Expression (33) as marginal pdfs and model parameter $\theta_2 = \lambda$ is chosen as model 2. Thus

$$f(y|m_1, \theta_1) = \frac{\theta_1 y_0^{\theta_1}}{y^{\theta_1+1}} \qquad \text{and} \qquad f(y|m_2, \theta_2) = \theta_2 \exp(-\theta_2(y - y_0)), \qquad y \geq y_0. \tag{48}$$

A common lower bound y_0 is used in the two models.

The model choice problem is based on observations $\boldsymbol{X} = (\boldsymbol{Z}, \boldsymbol{S})$ of maximum displacements obtained in two dimensions, see Section 5, where the number of observed faults is given as $n = \sum_{j=1}^{K} S_j$ and the observed maximum displacements are given as Z_j for all $\{j; s_j = 1\}$. The two competing models are assigned equal prior probabilities, $f(m_1) = f(m_2) = 0.5$, thus there is no prior belief that one model is more likely than the other.

The variable K represents the same quantity under both models. As mentioned above, auxiliary variables K_1 and K_2 are introduced to represent this variable under the two competing models in the McMC algorithm described by Borgos (2000). The main reason for this is that the posterior distribution of K can be bimodal, with one mode corresponding to each model, and auxiliary variables are introduced to avoid this bimodality. Depending on how prior information about the number of faults is obtained, it may also be appropriate to define different prior distributions of K under the two models, as discussed in Section 4.5. This is obtained by choosing individual intensities ω_m in the Poisson distribution in Expression (36).

6.2 Inference under the favored model

When a model choice has been made based on a set of observations, the next aim is to estimate the corresponding model parameters. Estimation of the power law exponent in a fractal model is frequently found in the literature, and is used in prediction of the number of faults in the observation area or in extrapolation to sub-seismic faults, see Gauthier and Lake (1993), Walsh et al. (1994). The standard estimation procedure is to fit a straight line to the log-log plot of the cumulative number of faults versus fault size. For discussions and improvements on this estimation technique, see Pickering et al. (1996). Pickering et al. (1995) and Clark et al. (1999) compared different procedures for estimating the exponent in a negative power law distribution, for example maximum likelihood estimation.

In this work, the model parameters in the favored model is estimated based on McMC simulations from the posterior distribution. Since the algorithm generates realizations of the model indicator and the model parameters simultaneously, parameter estimation can be performed based on output from the same run of the algorithm. An estimate of the parameter value in model m is given by the estimated posterior mean

$$\hat{\theta}_m = \frac{\sum_{j=1}^{r} \theta_m^j \mathbf{1}[m^j = m]}{\sum_{j=1}^{r} \mathbf{1}[m^j = m]}. \tag{49}$$

The number of faults in the observation area can be estimated similarly based on the realizations of K_m under the favored model, and the unknown fault sizes $\boldsymbol{Y}$ can be generated from the posterior distribution $f(\boldsymbol{y}|\boldsymbol{x}, m, k_m, \theta_m)$, see Borgos (2000).

The number of faults in the three dimensional reservoir, denoted $K_{m,3}$ under the favored model, can be predicted based on the estimate of the number of faults in two dimensions, now denoted $K_{m,2}$. The expected number in two dimensions, given $K_{m,3}$, is

$$\mathrm{E}\left[K_{m,2}|K_{m,3}\right] = f_{I_2}(1|m)K_{m,3}$$

where $f_{I_2}(1|m)$ is the probability of intersecting a fault by a two dimensional plane, under model m, see Expression (6). In a reservoir of vertical thickness T_v, the probability is given

as

$$f_{I_2}(1|m) = T_v^{-1}\mathrm{E}\left[h(D)|m\right] = \text{const} \times T_v^{-1}\mathrm{E}\left[D|m\right],$$

omitting the dip in the argument of the height function $h(\cdot)$, see Expression (3). The assumption $\tau = 1$ in Expression (23) is used, and the thickness of the reservoir is assumed to fulfill $T_v \gg h(D)$ for any maximum displacement D in the observation area, such that the vertical extent of all faults are completely contained in the region under study. The constant term depends on the distribution of the dip and on the constants c_1 and c_2, quantities which are not discussed in this work. When $K_{m,2}$ is observed or estimated, an estimate of the frequency $K_{m,3}/T_v$ is given by

$$\frac{\hat{K}_{m,3}}{T_v} = \frac{k_{m,2}}{f_{I_2}(1|m)T_v} = \text{const} \times \frac{k_{m,2}}{\mathrm{E}\left[D|m\right]}. \tag{50}$$

The mean value of D can be estimated for the chosen distribution, using the corresponding estimated model parameter $\hat{\theta}_m$. Under the fractal model, this is found to be

$$\hat{\mathrm{E}}\left[D|m_1\right] = \hat{\theta}_1^{-1}(\hat{\theta}_1 + 1)d_{03},$$

and in the exponential model

$$\hat{\mathrm{E}}\left[D|m_2\right] = \hat{\theta}_2^{-1} + d_{03}.$$

6.3 Exact observations

In this section the special case of no sampling errors is considered, where the exact number of faults is observed and fault sizes are observed without measurement errors. In this case the observations are $\boldsymbol{X} \equiv \boldsymbol{Y}$, and the likelihood function in Expression (41) is replaced by $h(\boldsymbol{x}|\boldsymbol{y}) = \delta_{\boldsymbol{x}}(\boldsymbol{y})$. The number of faults, K, is not treated as a stochastic variable in this case. The two proposed pdfs of Y under model 1 and 2, Expression (48), are now also the pdfs of X under the respective models. Bayes factor can in this situation be found analytically, using gamma priors $\mathrm{Gamma}(\alpha_m, \sigma_m)$ for the parameters θ_m, with mean $\mathrm{E}\left[\theta_m\right] = \alpha_m\sigma_m$ and variance $\mathrm{Var}\left[\theta_m\right] = \alpha_m\sigma_m^2$. The calculation of Bayes factor involves solving some integrals which are recognized as gamma functions, and is demonstrated by Borgos (2000). In this case Bayes factor is:

$$B_{12}(\boldsymbol{x}) = (\prod_{j=1}^{k} x_j)^{-1}\frac{\sigma_2^{\alpha_2}(\sigma_2^{-1} + \sum_{j=1}^{k}(x_j - x_0))^{\alpha_2+k}}{\sigma_1^{\alpha_1}(\sigma_1^{-1} + \sum_{j=1}^{k}(\ln x_j - \ln x_0))^{\alpha_1+k}} \cdot \frac{\Gamma(\alpha_2)\Gamma(\alpha_1+k)}{\Gamma(\alpha_1)\Gamma(\alpha_2+k)}. \tag{51}$$

The resulting posterior distributions of the parameters are gamma distributions with parameters $\alpha_m + k$ and $(\sigma_m^{-1} + t_m(\boldsymbol{x}))^{-1}$, where $t_1(\boldsymbol{x}) = \sum_{j=1}^{k}(\ln x_j - \ln x_0))$ and $t_2(\boldsymbol{x}) = \sum_{j=1}^{k}(x_j - x_0))$. The posterior mean and variance are given as

$$\mathrm{E}\left[\theta_m|\boldsymbol{x}\right] = \frac{\sigma_m(\alpha_m + k)}{1 + \sigma_m t_m(\boldsymbol{x})}, \qquad \mathrm{Var}\left[\theta_m|\boldsymbol{x}\right] = \frac{\sigma_m^2(\alpha_m + k)}{(1 + \sigma_m t_m(\boldsymbol{x}))^2}. \tag{52}$$

The model choice problem can thus be solved analytically, and the posterior mean can be used as estimators of the model parameters.

7 Example: Seismic Data

Figure 12 shows a dataset containing maximum displacements $\boldsymbol{x}$ observed on a two dimensional surface, interpreted from seismic data. The data are from the Gullfaks field in the North sea, see Fossen and Rørnes (1996). The number of observations is $n = 169$, with displacement values ranging from $2m$ to $256m$. Figure 12a shows a histogram of the observed displacements, corresponding to the pdf $f(x)$. Figure 12b shows a log-log plot of $\hat{S}(d)$, see Section 4.2 for an explanation.

The pdfs in Expression (48) are used as competing models, with $y_0 = 2m$ chosen as the minimum of the observations. The model parameters of the two competing models are assigned gamma priors with hyper-parameters $\alpha_1 = \alpha_2 = 4$, $\sigma_1 = 0.25$ and $\sigma_2 = 0.005$. Thus the prior mean values are $\mathrm{E}[\theta_1|m_1] = 1.0$ and $\mathrm{E}[\theta_2|m_2] = 0.02$. If the observations are assumed to be exact, Bayes factor from Expression (51) is given as $B_{12}(\boldsymbol{x}) = \mathcal{O}(10^{-35}) \approx 0$ and the posterior mean values are $\mathrm{E}[\theta_1|\boldsymbol{x}] = 0.393$ and $\mathrm{E}[\theta_2|\boldsymbol{x}] = 0.0232$, see Expression (52). In this case the exponential model is undoubtedly favored. However, the seismic data are not exact and the estimation of Bayes factor should account for the uncertainty caused by sampling errors. These errors are modeled by the likelihood function. Different choices of parameters in the likelihood function and prior distributions can lead to different conclusions of the model choice problem. This is demonstrated through a number of simulations examples, using the McMC algorithm described by Borgos (2000). In all simulations reported below, the estimates of model parameters and the number of faults are based on 5000 realizations from the McMC algorithm, separated by 10 iterations, while all 50 000 iterations are used to estimate Bayes factor. A sensitivity analysis is first carried out, exploring the influence of the parameters in the prior pdf and the likelihood function. Finally, well information is used to determine which prior parameters are most realistic.

Only a few parameters are varied in the simulations, the parameters η_2 and p_1 in the likelihood function and the prior mean numbers of faults ω_m under the two models. The likelihood function of the observed displacements $\boldsymbol{x} = (\boldsymbol{z}, \boldsymbol{s})$, Expression (41), contains the functions $p(y)$ given in Expression (42) and $h(z|y)$ given in (43). In the simulation study the function $h(z|y)$ is held constant, with parameter values $\delta_1 = 1m$, $\delta_2 = 5m$ and $\eta = \infty$. Thus the measurement error of each fault displacement is assumed to be in the range $1m$ to $5m$, while no censoring is assumed to occur. The influence of the parameters η_2 and p_1 in the function $p(y)$ is studied. The other values are held constant with values $\eta_1 = 2m$, chosen as the minimum value of the observations, and $p_2 = 0.99$. For $y \geq \eta_2$ there is a constant probability of 0.99 of observing the fault, while for decreasing $y \leq \eta_2$ the probability is decreasing to p_1 for $y = 2m$. Table 1 gives a summary of all parameter

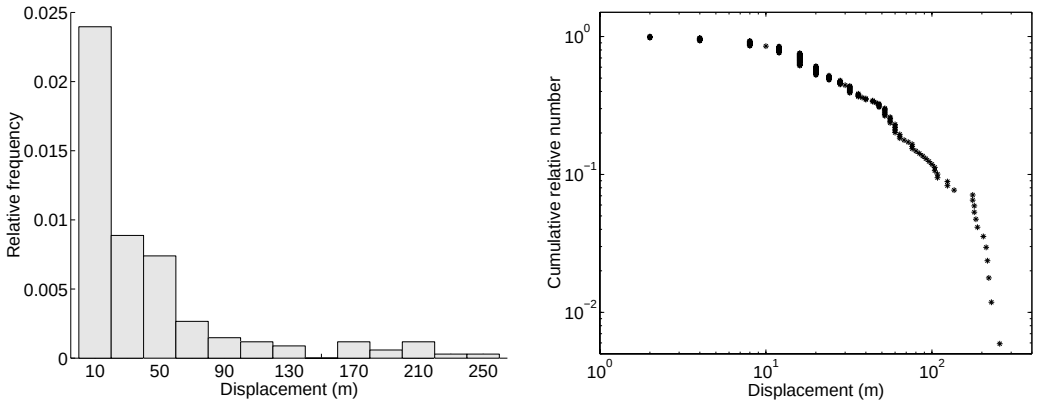


Figure 12: Maximum displacements observed on a two dimensional surface. (a) Histogram of the observations. (b) Log-log plot of $\hat{S}(d)$, see Expression (10).

values held constant in the simulations, and the values of the hyper-parameters in the prior distributions of the model parameters. Preliminary estimates of the number of faults in two dimensions based on well observations, see Section 4.5, show that for typical values of Θ_1 and Θ_2 the number of faults in two dimensions under the fractal model exceeds the number of faults under the exponential model. Thus throughout the simulation study, $\omega_1 \geq \omega_2$ is chosen. Assessment of ω_m based on well observations from the Gullfaks field is discussed in Section 7.4.

Figure 13 shows two realizations of $(\boldsymbol{Y}, K)$ from the posterior distribution (44) conditioned to the observations $\boldsymbol{x}$. For these realizations, $\omega_1 = \omega_2 = 720$, $p_1 = 0.1$ and $\eta_2 = 20m$. Figure 13a shows a realization from the fractal model, while Figure 13b shows a realization from the exponential model. The figures show log-log plots of the estimates $k\hat{S}(y)$ based on the realizations, see Expression (10), plotting absolute number instead of relative number. The estimates are compared with the estimate $n\hat{S}(z)$ for the observed fault sizes and the functions $kS(y)$ obtained using Expression (48) with estimated model parameters. In both figures a number of unobserved small faults are present in the realization of $\boldsymbol{Y}$, and the left part of the estimate $\hat{S}(y)$ deviates to some extent from the corresponding estimate $\hat{S}(z)$ based on the observations. In the right part of the figures there are only minor differences between the curves based on the realizations and the observations. In this area almost all faults are observed, and since $\eta = \infty$ no censoring is adjusted for in the realizations.

Prior distribution	Likelihood function					Hyper-Parameters			
y_0	δ_1	δ_2	η	p_2	η_1	α_1	σ_1	α_2	σ_2
2	1	5	∞	0.99	2	4	0.25	4	0.005

Table 1: Parameter values held constant in all simulations.

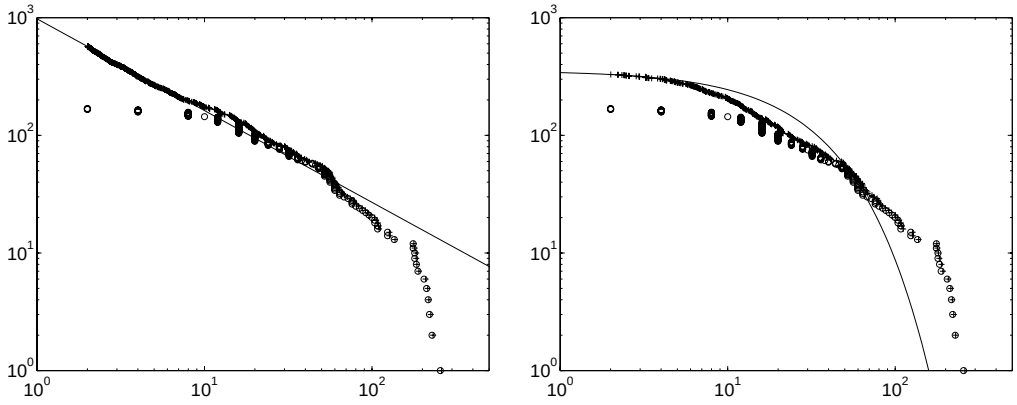


Figure 13: Realizations of $\boldsymbol{Y}$ from the posterior distribution $f(\boldsymbol{y}|\boldsymbol{x}, m, k_m, \theta_m)$ under (a) model 1 and (b) model 2. Estimates $\hat{\theta}_1 = 0.780$, $\bar{k}_1 = 570$, $\hat{\theta}_2 = 0.0370$ and $\bar{k}_2 = 329$ are obtained using parameters $\omega_1 = \omega_2 = 720$, $\eta_2 = 20m$ and $p_1 = 0.1$. The cumulative absolute number of faults, $k_m \hat{S}(y)$, is plotted on log-log scale (+), see Expression (10), and is compared to the observations $n\hat{S}(z)$ ($\circ$). The corresponding function $k_m S(y)$ for the prior distribution of Y is plotted with solid line, using the estimated model parameters.

7.1 Dependence on the prior distribution

Dependence of the model choice on the prior distribution is studied by varying ω_1 and ω_2, keeping all parameters of the likelihood function constant. Thus different results of the model choice problem is a consequence of the variations in ω_m, $m = 1, 2$. The parameter value $\eta_2 = 20m$ is used, while p_1 is held constant at the value $p_1 = 0.1$ or $p_1 = 0.01$. The remaining likelihood variables are given in Table 1. Table 2 shows the results of McMC simulations, reporting the estimated Bayes factor, $\hat{B}_{12}(\boldsymbol{x})$, for the observations in Figure 12, and the conclusion of the model choice problem. The table also lists the estimated model parameters, $\hat{\theta}_1$ and $\hat{\theta}_2$, and the estimated posterior mean number of faults, $\hat{\mathrm{E}}\left[K_m|\boldsymbol{x}, m\right] = \bar{K}_m$, for the favored model m.

From Table 2 it is clear that the prior mean numbers of faults have great influence on the result of the model choice problem. Especially, Table 2b illustrates how fast the conclusion of the model choice problem can change with changes in prior mean values. In most of the runs reported in the table, the exponential model is favored. However, with a relatively large, equal prior mean number of faults under both models, the fractal model is favored. From Table 2 it may seem that the model with the lowest prior mean number of faults in general is favored. However, comparing the posterior mean number of faults, the model with the higher mean number may be favored. In the simulation where $\omega_1 = \omega_2 = 720$ in Table 2b, no clear conclusion can be drawn. In this case, the posterior mean number of faults under model 1 is $\hat{k}_1 = 570$, while under model 2 it is $\hat{k}_2 = 329$, thus the posterior

distribution of the number of faults alone is not crucial to the model choice.

In Table 2a,b the value of p_1 is $p_1 = 0.1$, while in Table 2c it is decreased to $p_1 = 0.01$. With a lower probability of observing faults at the limit of resolution, the posterior mean number of faults increases. Furthermore, different values of p_1 may result in different conclusions of the model choice, as is observed in the case $\omega_1 = \omega_2 = 700$. The dependence of the model choice on p_1 is studied below.

As ω_2, the prior mean number of faults under model 2, increases, the posterior mean number also increases, see Table 2. A corresponding change in parameter value θ_2 is observed. Similarly, the posterior mean number of faults and the parameter θ_1 under model 1 is observed to increase when the prior mean number of faults, ω_1, increases.

7.2 Dependence on the likelihood function

Dependence on the likelihood function is studied by assigning different values to η_2 and p_1, keeping ω_1 and ω_2 constant. Thus different results of the model choice are now a consequence of the variations in the values of η_2 or p_1. Both parameters η_2 and p_1 have influence on the number of unobserved faults. Table 3 shows the results of McMC simulations where η_2 is varied, with $p_1 = 0.1$ constant. In Table 4 the parameter p_1 is varied, keeping $\eta_2 = 20m$ constant. Estimates of Bayes factor, the conclusion of the model choice and the model parameter of the favored model are reported.

As η_2 increases, the fault size required for an almost certain observation increases. The overall probability of observation decreases, increasing the number of unobserved faults. The results of Table 3 show tendencies of an increasing Bayes factor as η_2 starts increasing, thus model 1 becomes more favorable. However, as η_2 increases further, Bayes factor decreases again. This effect is observed using prior mean values $\omega_1 = \omega_2 = 700$ and $\omega_1 = \omega_2 = 720$. For the dataset studied here, the posterior probability of the fractal model being the right model seems to have a maximum for values of η_2 in the range 20–40. The case $\omega_1 = 600$ and $\omega_2 = 400$ is also considered, in which case Bayes factor is approximately equal to zero for all reported values of η_2.

The number of unobserved faults also increases as p_1 decreases, since a small value of p_1 indicates a large proportion of unobserved faults at small scales. The effect of p_1 is illustrated in Table 4, for some choices of ω_1 and ω_2. In the case $\omega_1 = \omega_2 = 700$, model 2 is favored when $p_1 = 0.1$. As p_1 decreases, the conclusion changes and model 1 is favored. The prior mean values $\omega_1 = 600$ and $\omega_2 = 400$ are also considered. In this case model 2 is favored for all values of p_1 considered, and a reduction of p_1 from 0.1 to 0.0001 does not give a change of the conclusion.

ω_1	ω_2	$\hat{B}_{12}(\boldsymbol{x}, n)$	Model	$\hat{\theta}_1$	$\bar{k}_1$	$\hat{\theta}_2$	$\bar{k}_2$
400	400	0^*	2			0.0281	234
500	500	0^*	2			0.0303	256
600	600	$\mathcal{O}(10^{-17})$	2			0.0329	284
700	700	0.00538	2			0.0364	322
800	800	$\mathcal{O}(10^{12})$	1	0.822	624		
900	900	∞^*	1	0.874	702		
600	400	0^*	2			0.0283	235
700	500	0^*	2			0.0306	260
800	400	0^*	2			0.0283	235
800	600	0^*	2			0.0334	288
900	500	0^*	2			0.0311	265
900	700	0^*	2			0.0362	320
900	800	$3.25 \cdot 10^{-5}$	2			0.0387	349

(a)

ω_1	ω_2	$\hat{B}_{12}(\boldsymbol{x}, n)$	Model	$\hat{\theta}_1$	$\bar{k}_1$	$\hat{\theta}_2$	$\bar{k}_2$
700	700	0.00538	2			0.0364	322
710	710	0.115	2			0.0367	324
720	720	1.239	1/2	0.780	570	0.0370	329
730	730	169.8	1	0.770	559		

(b)

ω_1	ω_2	$\hat{B}_{12}(\boldsymbol{x}, n)$	Model	$\hat{\theta}_1$	$\bar{k}_1$	$\hat{\theta}_2$	$\bar{k}_2$
500	500	0^*	2			0.0317	271
700	500	0^*	2			0.0320	273
700	700	∞^*	1	0.795	589		
900	500	0^*	2			0.0322	273
900	700	0^*	2			0.0383	340
900	900	∞^*	1	0.915	773		

(c)

Table 2: Results of McMC simulations where $\eta_2 = 20m$ while various prior mean numbers ω_m are used. In (a) and (b) $p_1 = 0.1$ and in (c) this probability is reduced to $p_1 = 0.01$. Table (b) illustrates how the conclusion of the model choice can change for only minor changes of ω_1 and ω_2. The estimates of Bayes factor and the model parameters are obtained using Expressions (47) and (49). The estimates of the model parameters and the mean values $\bar{k}_m$ are based on 5000 realizations, separated by 10 iterations of the McMC algorithm. Bayes factor is estimated based on all 50 000 iterations. The reported values 0^* and ∞^* indicate that the estimate of Bayes factor is either very small or very large, without being identically 0 or ∞. $\mathcal{O}(u)$ gives the order of magnitude.

ω_1	ω_2	η_2	$\hat{B}_{12}(\boldsymbol{x}, n)$	Model	$\hat{\theta}_1$	$\bar{k}_1$	$\hat{\theta}_2$	$\bar{k}_2$
700	700	10	0^*	2			0.0275	225
700	700	20	0.00538	2			0.0364	322
700	700	40	0.129	2			0.0423	452
700	700	60	$\mathcal{O}(10^{-13})$	2			0.0420	527
700	700	80	0^*	2			0.0402	568
700	700	100	0^*	2			0.0382	585
720	720	10	0^*	2			0.0282	231
720	720	20	1.239	1/2	0.780	570	0.0370	329
720	720	30	$8.41 \cdot 10^4$	1	0.744	574		
720	720	40	0.743	1/2	0.743	621	0.0431	469
720	720	50	$5.25 \cdot 10^{-7}$	2			0.0430	507
600	400	20	0^*	2			0.0283	235
600	400	40	0^*	2			0.0320	295
600	400	60	0^*	2			0.0330	341
600	400	80	0^*	2			0.0326	369
600	400	100	0^*	2			0.0316	381

Table 3: Results of McMC simulations with $p_1 = 0.1$, while various parameter values of ω_1, ω_2 and η_2 are used. The estimates of Bayes factor and the model parameters are obtained using Expressions (47) and (49). The estimates of the model parameters and the mean values $\bar{k}_m$ are based on 5000 realizations, separated by 10 iterations of the McMC algorithm. Bayes factor is estimated based on all 50 000 iterations. The reported value 0^* indicates that the estimate of Bayes factor is very small, without being identically 0. $\mathcal{O}(u)$ gives the order of magnitude.

ω_1	ω_2	p_1	$\hat{B}_{12}(\boldsymbol{x}, n)$	Model	$\hat{\theta}_1$	$\bar{k}_1$	$\hat{\theta}_2$	$\bar{k}_2$
700	700	0.1	0.00538	2			0.0364	322
700	700	0.01	$2.40 \cdot 10^8$	1	0.791	583		
700	700	0.001	$9.00 \cdot 10^8$	1	0.805	601		
700	700	0.0001	$1.34 \cdot 10^9$	1	0.799	594		
600	400	0.1	0^*	2			0.0283	235
600	400	0.01	0^*	2			0.0292	245
600	400	0.001	0^*	2			0.0291	243
600	400	0.0001	0^*	2			0.0296	249

Table 4: Results of McMC simulations with $\eta_2 = 20m$, while various parameter values of ω_1, ω_2 and p_1 are used. The estimates of Bayes factor and the model parameters are obtained using Expressions (47) and (49). The estimates of the model parameters and the mean values $\bar{k}_m$ are based on 5000 realizations, separated by 10 iterations of the McMC algorithm. Bayes factor is estimated based on all 50 000 iterations. The reported value 0^* indicates that the estimate of Bayes factor is very small, without being identically 0.

7.3 Parameter estimation and prediction of number of faults

Tables 2–4 show results of McMC simulations for observations with sampling errors. Both Bayes factor and the model parameter under the favored model are estimated, see Expressions (47) and (49). Figure 14 shows histograms of realizations of θ_1 and θ_2, for some of the simulation examples from Table 2. The histograms are compared with the corresponding prior pdf, with hyper parameters given in Table 1. The histograms give an estimate of the posterior pdf, showing how the modes of the posterior pdfs differ from the modes of the prior pdfs. Furthermore, the histograms have narrower ranges than the prior pdfs, thus the posterior distributions have a smaller variance than the prior distributions. As ω_m increases, the mode of the posterior pdf $f(\theta_m|\boldsymbol{x})$ also increases. This coincides with the results of Table 2, where the posterior mean is observed to increase with ω_m.

Since the relationship between model parameters in three and two dimensions is known both for the fractal model and the exponential model, see Section 4, the parameters in three dimensions can be estimated based on the corresponding estimates in two dimensions. Thus, in cases where the fractal model is favored, the power law exponent in the displacement distribution in three dimensions is estimated by $\hat{\beta}_3 = \hat{\beta}_2 + 1 = \hat{\theta}_1 + 1$, under the assumption $\tau = 1$, see Expression (23). In the exponential model the distribution of displacement has the same parameter in all dimensions, with estimate $\hat{\lambda} = \hat{\theta}_2$.

When a proper model is chosen for the distribution of displacement in three dimensions, the final goal is to predict the number of faults in a three dimensional reservoir. The procedure is described in Section 6.2. To illustrate the consequence of the model choice on the predicted number of faults in three dimensions, the estimates of the model parameters Θ_m in Table 2b are used, with $\omega_1 = \omega_2 = 720$. This situation gives no conclusion to the model choice problem. The estimates of model parameters and fault numbers in two dimensions are $\hat{\theta}_1 = 0.780$, $\bar{k}_{1,2} = 570$, $\hat{\theta}_2 = 0.0370$ and $\bar{k}_{2,2} = 329$. The constant term in Expression (50) depends on the constants c_1 and c_2 from Expression (1) and the distribution of the dip. A typical dip of 60° is assumed, and c_1 and c_2 are chosen inspired by the work of Walsh and Watterson (1989) and Dawers et al. (1993). The result is a constant term of value 0.035. By the use of Expression (50), the frequencies of faults in three dimensions of displacement above $d_{03} = 2m$, vertically through the reservoir, are found to be $\hat{k}_{1,3} = 4370\,km^{-1}$ and $\hat{k}_{2,3} = 397\,km^{-1}$ under model 1 and 2 respectively. Thus in this case the estimates under the two models differ by a factor 11, illustrating the possible consequences of choosing a wrong model. No boundary effects are accounted for in these calculations.

7.4 Most favorable model based on well information

Hesthammer and Fossen (1997) studied fault intensities in wells in the Gullfaks fields, and found an estimate of 4.4 faults per km in the east-west direction, perpendicular to the

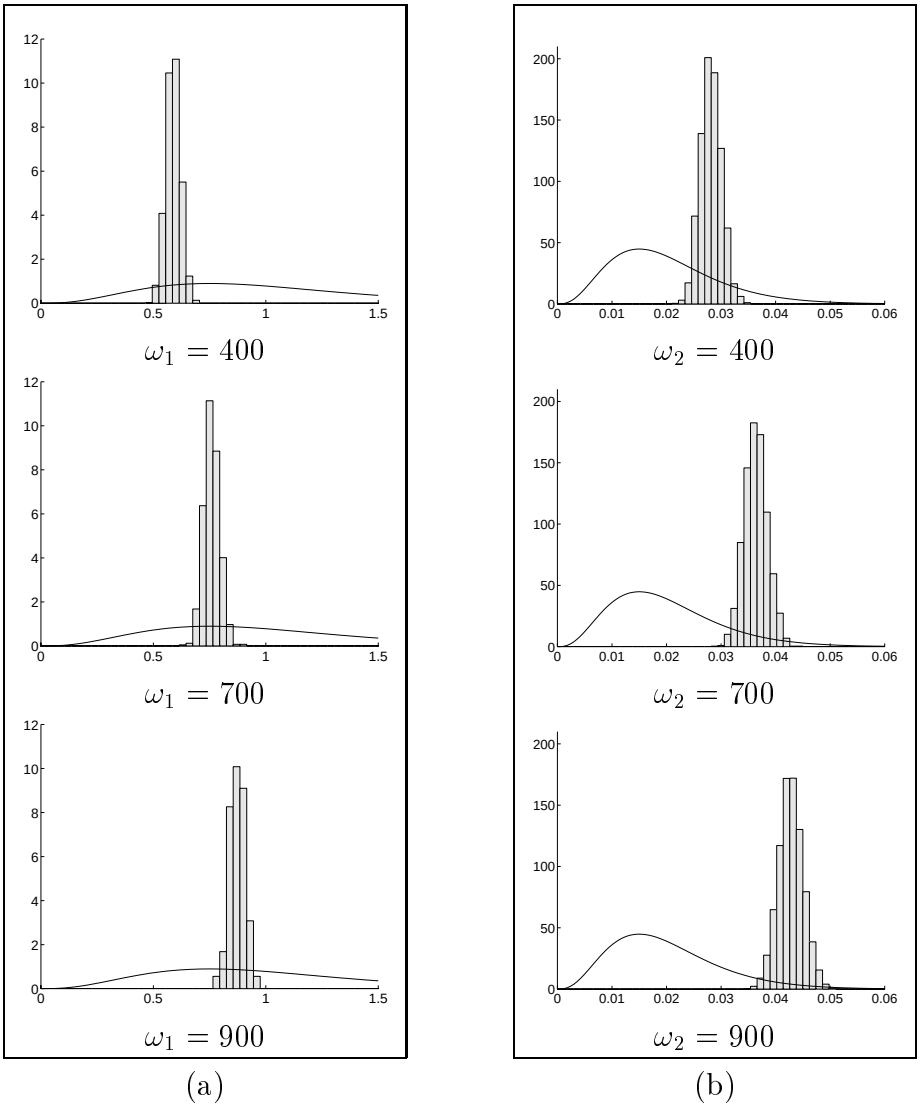


Figure 14: (a) Histograms of realizations of the model parameter θ_1 in model 1 from $f(\theta_1|\boldsymbol{x}, m_1)$, obtained from McMC simulations. The solid lines show the prior pdf $f_1(\theta_1|m_1)$. (b) Similar histograms of realizations of θ_2 from $f(\theta_2|\boldsymbol{x}, m_2)$, comparing with $f_2(\theta_2|m_2)$. The values of the prior means ω_1 and ω_2 are given below each figure, and $\eta_2 = 20m$ and $p_1 = 0.1$.

dominating strike direction. This number only includes faults of displacement larger than approximately $10m$ and gives a lower limit of fault intensity, since also well observations can suffer from missing observations. This fault intensity can be converted into a number of faults in the two dimensional observation area, which can serve as a prior mean number. The procedure is described in Section 4.5. The simulations above show that the model parameters Θ_1 and Θ_2 typically lies in the intervals $\Theta_1 \in [0.8, 0.9]$ and $\Theta_2 \in [0.025, 0.04]$. Since Θ_1 seems to be less than 1, the mean value $\mathrm{E}_2[D|m_1]$ needed in Expression (37) does not exist. To deal with this problem, $\mathrm{E}_2[D|m]$ is replaced by $\mathrm{E}_2[D|m, D \leq d_{\max}]$, where $d_{\max}$ is chosen larger than the maximum observed displacement value. Assuming a value $c_1 = 75$ of the constant involved in Expressions (1) and (37), and using a rough approximation of $8km \times 12.8km$ for the observation area, see Fossen and Rørnes (1996), the number of faults of displacement above $2m$ are found to be 450–600 under model 1 and 150–250 under model 2. These numbers give an indication of how large the prior mean number of faults under the two models at least should be chosen.

The sensitivity analysis presented above mainly concentrate on prior mean values ω_1 and ω_2 equal to or larger than the values found based on well information. Tables 2–4 show that the fractal model is only considered best in cases where the prior mean values are $\omega_m > 700$ for both m. Such a large value can be realistic for ω_1, but is clearly too large for ω_2 based on the well information. Among the simulations presented above, the ones with prior mean values $\omega_1 = 600$ and $\omega_2 = 400$ are more realistic. In this case the exponential model is clearly most favorable, regardless of the likelihood parameters included in the simulation study.

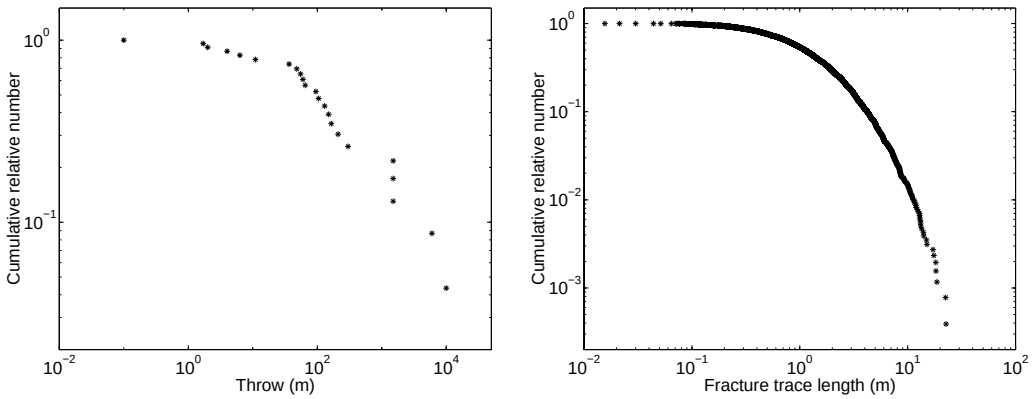
8 Example: Outcrop Data

Figure 15 shows log-log plots of faults observed in outcrops. Figure 15a shows $n = 23$ throw values obtained in a line sample in western Sinai, see Knott et al. (1996) and Beach et al. (1999). Throw is measured as the vertical component of displacement. Figure 15b shows $n = 2565$ fracture trace lengths observed in a two dimensional observation area in Hornelen, Norway, see Odling (1997). In contrast to faults, fractures have no displacement of rock.

The two models in Expression (48) are used as competing models for the dataset from western Sinai, where θ_1 and θ_2 are now the parameters of the distributions of displacements in one dimension, see Expressions (17) and (33). The common lower limit y_0 is chosen equal to the minimum observation. If the line sample observations from western Sinai are assumed to be exact, with no measurement errors or missing observations, maximum likelihood estimates of the parameters under the two models are given as $\hat{\theta}_1 = 0.151$ and $\hat{\theta}_2 = 0.00105$. Using these estimates as prior means in gamma distributions with shape parameters $\alpha_m = 4$, gives a Bayes factor of $B_{12}(\boldsymbol{x}) = 3.04 \cdot 10^6$. Other reasonable choices

of prior mean values also give a large value of Bayes factor, and the fractal model is undoubtedly favored compared to the exponential model.

The observed fracture trace lengths from Hornelen are also assumed to be exact. The distributions of fault width in two dimensions are compared for this dataset, given in Expression (20) for the fractal model and in Expression (30) or (34) under the exponential model. Neither of the distributions from the exponential model give analytically tractable expressions of Bayes factor. If a first order approximation of the modified Bessel function of order 0 is used, see Expression (58), the distribution in Expression (30) can be approximated by a gamma distribution $[Y_j|m_2, \theta_2] \sim \text{Gamma}(1.5, \theta_2)$. In this case Bayes factor is analytically tractable, and reasonable choices of hyper-parameters in the prior pdfs of the model parameters give $B_{12}(\boldsymbol{x}) \approx 0$. Thus in this case the exponential model is far better than the fractal model. Maximum likelihood parameter estimation gives $\hat{\theta}_1 = 0.236$ and $\hat{\theta_2} = 0.815$. The dataset is one of several datasets gathered at different scales, and Odling (1997) concludes with a log-normal distribution for each single dataset, while a fractal model is found to describe the coalescent fault population. The log-normal distribution is not considered as an alternative distribution in the model choice problem described in the current work, but the same approach as illustrated in this work could be applied to compare this distribution with either the exponential or the fractal model.



(a) (b)

Figure 15: (a) Fault displacements observed in one dimension in western Sinai. (b) Fracture trace lengths observed in two dimensions in Hornelen, Norway.

9 Conclusion

A proper model should be chosen and the model parameters estimated, in order to make inference about a fault population in three dimensions. The estimated number of faults and their sizes depend on which model is used to obtain the estimates. Observation

errors should be accounted for in the model choice, and the relationship between the three dimensional fault population and the observations obtained in lower dimensions must be known.

Forward calculations are performed to derive the relationship between the distributions of maximum displacement in three, two and one dimensions, starting with both a fractal and an exponential distribution in three dimensions. Corresponding distributions of fault width are obtained using the relationship $W \propto D^{1/\tau}$. Using some basic assumptions on fault geometry and spatial distribution, it is shown that if the distribution of displacement in three dimensions is fractal, then so is the distribution of fault width and the corresponding distributions in two and one dimensions. The decrease in the power law exponent for the displacement both from three to two dimensions and from two to one is $1/\tau$, while for fault width the power law exponent decreases by 1 from three to two dimensions. The relationships found between power law exponents of displacement or fault width in one, two and three dimensions correspond to previously published results, see Heffer and Bevan (1990), Marrett and Allmendinger (1991), Piggott (1997) and Berkowitz and Adler (1998).

When an exponential distribution of displacement in three dimensions is assumed, the derived expressions of the other distributions are less analytically tractable than the corresponding ones under the fractal model. The expressions simplifies if $\tau = \gamma = 1$, where a linear relationship between width and displacement is assumed and a linear displacement function is applied, see Expressions (1) and (5). In this case the displacement follows an exponential distribution with the same parameter in three, two and one dimensions, thus observations of exponentially distributed displacements do not suffer from over-representation of large faults in lower dimensions. Under the assumption $\tau = 1$ the distribution of the width in three dimensions is also exponential, while in two dimensions the distribution is not exponential.

Section 6 describes a statistical approach for choosing between two suggested models for fault size distributions. The example in Section 7 shows how the conclusion of the model choice depends on the parameters both in the prior distribution of the number of faults and on the likelihood function. Parts of the parameter space is spanned to illustrate how the conclusion varies as the prior model or the assumptions on the sampling errors change. The prior distribution and likelihood function thus have great influence on the model choice, and in order to get the right conclusion the parameters of the prior pdf and the likelihood model should be chosen based on general geological experience and knowledge about the reservoir and the data collection.

If there is a strong belief in a negative power law distribution of fault sizes, based solely on prior experience, it can be tempting to define the prior and likelihood parameters based on the observations. For example, η_2 can be chosen as a break in the log-log curve and ω_1 can be chosen by extrapolating a line fitted to data with $d \geq \eta_2$ to find the total number of faults. This may lead to an optimal choice of parameter values, in the sense that the fractal model is more likely to be favored. However, this procedure uses the observations to

define both the prior distribution and the likelihood function, thus reducing the reliability of the statistical analysis.

The prior mean numbers of faults under the two models should be related to the lithology and the tectonic history of the reservoir under study, and can also be calibrated with fault observations from wells. The likelihood functions should be related to the data acquisition. For example, for seismic data the likelihood parameters are related to the seismic resolution, and thus the seismic wavelength. The interpretation process can also have influence on the likelihood parameters, as the detection of small faults can depend both on the interpreter and on the time spent on the interpretation. If specific values for the likelihood parameters are hard to define, then alternatively the parameters can be assigned prior distributions. The statistical analysis presented illustrates the importance of choosing the prior and likelihood parameters correctly. For a given dataset from the Gullfaks field, it is shown how the fault intensity observed in wells is used to determine realistic prior mean number of faults in the observation area under both competing models, and based on these values the exponential model is found to describe the fault displacements displacements better than the fractal model.

References

Abramowitz, M. and Stegun, I. A. (eds) (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Dover books on advanced mathematics, Dover, New York.

Ackermann, R. V., Withjack, M. O. and Schlische, R. W. (1999). The geometric and statistical evolution of normal fault systems: An experimental study of the effects of mechanical layer thickness on scaling laws, submitted to *Journal of Structural Geology*.

Badley, M. E., Freeman, B., Roberts, A. M., Thatcher, J. S., Walsh, J., Watterson, J. and Yielding, G. (1990). Fault Interpretation During Seismic Interpretation and Reservoir Evaluation, *The Integration of Geology, Geophysics, Petrophysics and Petroleum Engineering in Reservoir Delineation, Description and Management*, Proceedings of the 1st Archie Conference, Houston, Texas, pp. 224–241. AAPG.

Barnett, J. A. M., Mortimer, J., Rippon, J. H., Walsh, J. J. and Watterson, J. (1987). Displacement Geometry in the Volume Containing a Single Normal Fault, *The American Association of Petroleum Geologists Bulletin* **71**(8): 925–937.

Beach, A., Welbon, A. I., Brockbank, P. J. and McCallum, J. E. (1999). Reservoir damage around faults: outcrop from the Suez rift, *Petroleum Geoscience* **5**: 109–116.

Berkowitz, B. and Adler, P. M. (1998). Stereological analysis of fracture network structure in geological formations, *Journal of Geophysical Research* **103**(B7): 15339–15360.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion), *Statistical Science* **10**(1): 3–66.

Borgos, H. G. (2000). Sampling algorithm for estimating Bayes factor, Part II of this thesis.

Carlin, B. P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods, *J. R. Statist. Soc. B* **57**(3): 473–484.

Childs, C., Walsh, J. J. and Watterson, J. (1990). A Method for Estimation of the Density of Fault Displacements below the Limits of Seismic Resolution in Reservoir Formations, *North Sea Oil and Gas Reservoirs - II*, Graham & Trotman, pp. 309–318.

Chilès, J. P. (1988). Fractal and Geostatistical Methods for Modeling of a Fracture Network, *Mathematical Geology* **20**(6): 631–654.

Clark, R. M., Cox, S. J. D. and Laslett, G. M. (1999). Generalization of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats, *Geophysical Journal International* **136**: 357–372.

Cowie, P. A. (1998). Normal Fault Growth in Three-Dimensions in Continental and Oceanic Crust, *Faulting and Magmatism at Mid-Ocean Ridges*, Geophysical Monograph 106, the American Geophysical Union, pp. 325–348.

Cowie, P. A. and Scholz, C. H. (1992a). Physical explanation for the displacement-length relationship of faults using a post-yield fracture mechanics model, *Journal of Structural Geology* **14**(10): 1133–1148.

Cowie, P. A. and Scholz, C. H. (1992b). Displacement-length scaling relationship for faults: data synthesis and discussion, *Journal of Structural Geology* **14**(10): 1149–1156.

Cowie, P. A., Malinverno, A., Ryan, W. B. F. and Edwards, M. H. (1994). Quantitative fault studies on the East Pacific Rise: A comparison of sonar image techniques, *Journal of Geophysical Research* **99**(B8): 15205–15218.

Dawers, N. H., Anders, M. H. and Scholz, C. H. (1993). Growth of normal faults: Displacement-length scaling, *Geology* **21**: 1107–1110.

Dawers, N. H. and Anders, M. H. (1995). Displacement-length scaling and fault linkage, *Journal of Structural Geology* **17**(5): 607–614.

Evans, M. and Swartz, T. (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems, *Statistical Science* **10**(3): 254–272.

Feder, J. (1988). *Fractals*, Plenum Press, New York.

Fossen, H. and Rørnes, A. (1996). Properties of fault populations in the Gullfaks Field, northern North Sea, *Journal of Structural Geology* **18**(2/3): 179–190.

Gauthier, B. D. M. and Lake, S. D. (1993). Probabilistic Modeling of Faults Below the Limit of Seismic Resolution in Pelican Field, North Sea, Offshore United Kingdom, *The American Association of Petroleum Geologists Bulletin* **77**(5): 761–777.

Han, C. and Carlin, B. P. (2000). MCMC Methods for Computing Bayes Factors: A Comparative Review, *Research Report 2000-001*, Division of Biostatistics, University of Minnesota.

Hatcher, R. D., Jr. (1995). *Structural Geology. Principles, Concepts, and Problems*, 2 edn, Prentice Hall, Englewood Cliffs, New Jersey.

Heffer, K. J. and Bevan, T. G. (1990). Scaling Relationships in Natural Fractures: Data, Theory, and Application, *Europec 90*, Society of Petroleum Engineers, pp. 367–376. SPE Paper 20981.

Hesthammer, J. and Fossen, H. (1997). The influence of seismic noise in structural interpretation of seismic attribute maps, *First Break* **15**(6): 209–219. (C. Townsend, pers.comm.).

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2 edn, Wiley.

Jones, G., Fisher, Q. J. and Knipe, R. J. (eds) (1998). *Faulting, Fault Sealing and Fluid Flow in Hydrocarbon Reservoirs*, number 147 in *Special Publications*, Geological Society, London.

JSG (1996). *Journal of Structural Geology*, Vol. 18, No. 2/3, Special Issue, Scaling Laws for Fault and Fracture Populations – Analyses and Applications.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association* **90**(430): 773–795.

Knott, S. D., Beach, A., Brockbank, P. J., Brown, J. L., McCallum, J. E. and Welbon, A. I. (1996). Spatial and mechanical controls on normal fault populations, *Journal of Structural Geology* **18**(2/3): 359–372.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York.

Marrett, R. and Allmendinger, R. W. (1991). Estimates of strain due to brittle faulting: sampling of fault populations, *Journal of Structural Geology* **13**(6): 735–738.

Marrett, R. and Allmendinger, R. W. (1992). Amount of extension on "small" faults: An example from the Viking graben, *Geology* **20**: 47–50.

Munthe, K., Holden, L., Mostad, P. and Townsend, C. (1994). Modelling Sub-seismic Fault Patterns Using a Marked Point Process, *Proceedings of the 4th European conference on the Mathematics of Oil Recovery, Røros, Norway.* Topic B: Heterogeneity Description and Assessment of Uncertainty.

Munthe, K. L., Omre, H., Holden, L., Damsleth, E., Heffer, K., Olsen, T. S. and Watterson, J. (1993). Subseismic Faults in Reservoir Description and Simulation, *68th Annual Conference and Exhibition of the Society of Petroleum Engineers*, Society of Petroleum Engineers, pp. 843–850. SPE Paper 26500.

Oakeshott, R. B. S. and Edwards, S. F. (1992). On the stereology of ellipsoids and cylinders, *Physica A* **189**(1/2): 208–233.

Odling, N. E. (1997). Scaling and connectivity of joint systems in sandstones from western norway, *Journal of Structural Geology* **19**(10): 1257–1271.

Pickering, G., Bull, J. M. and Sanderson, D. J. (1995). Sampling power-law distributions, *Tectonophysics* **248**: 1–20.

Pickering, G., Bull, J. M. and Sanderson, D. J. (1996). Scaling of fault displacements and implications for the estimation of sub-seismic strain, *in* P. C. Buchanan and D. A. Nieuwland (eds), *Modern Developments in Structural Interpretation, Validation and Modeling*, Geological Society Special Publication No. 99, pp. 11–26.

Piggott, A. R. (1997). Fractal relations for the diameter and trace length of disc-shaped fractures, *Journal of Geophysical Research* **102**(B8): 18121–18125.

Sassi, W., Livera, S. E. and Caline, B. P. R. (1992). Reservoir compartmentation by faults in Cormorant Block IV, U.K. northern North Sea, *in* R. M. Larsen, H. Brekke, B. T. Larsen and E. Talleraas (eds), *Structural and Tectonic Modelling and its Application to Petroleum Geology*, NPF Special Publication 1, Elsevier, Amsterdam, pp. 355–364.

Spryopoulos, C., Griffith, W. J., Scholz, C. H. and Shaw, B. E. (1999). Experimental evidence for different strain regimes of crack populations in a clay model, *Geophysical Research Letters* **26**(8): 1081–1084.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, New York.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2 edn, Wiley, New York.

Walsh, J. J. and Watterson, J. (1987). Distribution of cumulative displacement and seismic slip on a single normal fault surface, *Journal of Structural Geology* **9**(8): 1039–1046.

Walsh, J. J. and Watterson, J. (1988). Analysis of the relationship between displacements and dimensions of faults, *Journal of Structural Geology* **10**(3): 239–247.

Walsh, J. J. and Watterson, J. (1989). Displacement gradients on fault surfaces, *Journal of Structural Geology* **11**(3): 307–316.

Walsh, J. J., Watterson, J. and Yielding, G. (1991). The importance of small-scale faulting in regional extension, *Nature* **351**: 391–393.

Walsh, J. J., Watterson, J. and Yielding, G. (1994). Determination and Interpretation of Fault Size Populations: Procedures and Problems, *North Sea Oil and Gas Reservoirs - III*, Norwegian Institute of Technology, Kluwer Academic Publishers, pp. 141–155.

Watterson, J. (1986). Fault dimensions, displacement and growth, *Pure and Applied Geophysics* **124**: 365–373.

Yielding, G., Freeman, B. and Needham, D. T. (1997). Quantitative fault seal prediction, *The American Association of Petroleum Geologists Bulletin* **81**(6): 897–917.

A Approximation of Fault Heights

The height of a fault ellipsoid can be expressed as a function $h(x) = Kx^{-1}\sqrt{1+ax^2}$, omitting the dependence on the dip θ, with

$$K = \sin\theta c_2^{-1}, \quad a = \left(\frac{c_2 c_3}{\tan\theta}\right)^2, \quad x = w^{-1} = (dc_1)^{-1/\tau}. \tag{53}$$

The function $h(x)$ can further be expressed as $h(x) = Kx^{-1}g(x)$, where $g(x) = \sqrt{1+ax^2}$. A Taylor expansion of $g(x)$ gives

$$g(x) = 1 + \frac{ax^2}{2(1+a\xi^2)^{3/2}} = 1 + R(x), \qquad \text{for some } \xi \in [0, x].$$

Since $a\xi^2 \geq 0$ it follows that $R(x) \leq ax^2/2$. Let

$$\hat{h}(x) = Kx^{-1} = \sin\theta(xc_2)^{-1}.$$

Substituting w or $(dc_1)^{1/\tau}$ for x, Expressions (2) and (3) are obtained respectively. The error term of the approximation is $h(x) - \hat{h}(x) = Kx^{-1}g(x) - Kx^{-1} = Kx^{-1}R(x)$, and the absolute error has an upper bound

$$\frac{h(x)-\hat{h}(x)}{h(x)} = \frac{Kx^{-1}R(x)}{Kx^{-1}g(x)} \leq \frac{ax^2}{2\sqrt{1+ax^2}} = \left(\frac{c_2c_3}{\tan\theta}\right)^2 \frac{x^2}{2}\left(1+\left(\frac{c_2c_3}{\tan\theta}\right)^2 x^2\right)^{-1/2}.$$

This function can be shown to be monotonely increasing, thus substituting w or d from Expression (53), the relative error decreases monotonely for increasing fault size.

B Fault Distributions

B.1 Fractal distributions: displacement

The pdf of the maximum displacement D in three dimensions given in Expression (14) is:

$$f_3(d) = \frac{\beta_3 d_{03}^{\beta_3}}{d^{\beta_3+1}}, \qquad d \geq d_{03}.$$

Let D' be the displacement obtained in two dimensions. Using the approximation of the height of the fault ellipsoid given in (3) and the displacement function (5), the pdfs corresponding to Expressions (7) and (8) are found to be:

$$\begin{aligned} f(d|1) &= \frac{(\beta_3 - 1/\tau) d_{03}^{\beta_3 - 1/\tau}}{d^{\beta_3 - 1/\tau + 1}} \\ f(d'|d,1) &= \frac{(d')^{1/\gamma - 1}}{\gamma d^{1/\gamma}}, \quad 0 \leq d' \leq d \\ f(d'|1) &= \int_{d'}^{\infty} \frac{(d')^{1/\gamma-1}}{\gamma d^{1/\gamma}} \cdot \frac{(\beta_3 - 1/\tau) d_{03}^{\beta_3 - 1/\tau}}{d^{\beta_3 - 1/\tau + 1}} \, dd \\ &= \frac{\beta_3 - 1/\tau}{\gamma(\beta_3 - 1/\tau) + 1} \cdot \frac{d_{03}^{\beta_3 - 1/\tau}}{(d')^{\beta_3 - 1/\tau + 1}}, \quad d' \geq d_{03}. \end{aligned} \tag{54}$$

The last expression is equal to the pdf in Expression (15), omitting conditioning on $I = 1$.

B.2 Fractal distributions: width

The pdf of the width W in three dimensions given in Expression (18) is

$$f_3(w) = \frac{\mu_3 w_{03}^{\mu_3}}{w^{\mu_3+1}}, \qquad w \geq w_{03}.$$

Let W' be the fault width obtained in two dimensions. Using the approximation of the height of the fault ellipsoid given in (2) and the relationship (4) between w' and w at a given position of intersection, the pdfs corresponding to Expressions (7) and (8) are found to be:

$$f(w|1) = \frac{(\mu_3 - 1) w_{03}^{\mu_3 - 1}}{w^{\mu_3}}$$

$$f(w'|w,1) = \frac{w'}{w\sqrt{w^2 - (w')^2}} \tag{55}$$

$$f(w'|1) = \int_{w'}^{\infty} \frac{w'}{w\sqrt{w^2 - (w')^2}} \cdot \frac{(\mu_3 - 1) w_{03}^{\mu_3 - 1}}{w^{\mu_3}} \, dw. \tag{56}$$

The latter integral can be solved using the substitution $w = w'/\cos u$, which gives:

$$f(w'|1) = \frac{w'(\mu_3 - 1)w_{03}^{\mu_3 - 1}}{(w')^{\mu_3 + 1}} \int_0^{\pi/2} (\cos u)^{\mu_3}\, du = \frac{w_{03}^{\mu_3 - 1}(\mu_3 - 1)}{(w')^{\mu_3}} \cdot \frac{\Gamma((\mu_3 + 1)/2)}{\Gamma((\mu_3 + 2)/2)} \cdot \frac{\sqrt{\pi}}{2}.$$

This is the pdf in Expression (19), omitting conditioning on $I = 1$.

B.3 Exponential distributions: displacement

The pdf of the maximum displacement D in three dimensions given in Expression (24) is

$$f_3(d) = \lambda \exp(-\lambda d), \qquad d \geq 0.$$

Let D' be the displacement obtained in two dimensions. Using the approximation of the height of the fault ellipsoid given in (3), the pdf corresponding to Expression (7) is found to be:

$$f(d|1) = \frac{\lambda^{1+1/\tau}}{\Gamma(1 + 1/\tau)} d^{1/\tau} \exp(-\lambda d).$$

This is the pdf of a gamma distribution. The function $f(d'|d, 1)$ is the same as in the fractal case, given in Expression (54). The integral in Expression (8) gives the pdf of the displacement in two dimensions:

$$f(d'|1) = \gamma^{-1} \lambda^{1/\gamma} (d')^{1/\gamma - 1} \frac{\Gamma(1/\tau - 1/\gamma + 1, \lambda d')}{\Gamma(1 + 1/\tau)}, \qquad \gamma > \frac{\tau}{1 + \tau}.$$

Substituting D' by D and $f(d')$ by $f_2(d)$ and omitting conditioning on $I = 1$, Expression (25) is obtained.

B.4 Exponential distributions: width

The pdf of the width W in three dimensions given in Expression (28) is

$$f_3(w) = \rho\tau w^{\tau - 1} \exp(-\rho w^\tau), \qquad w \geq 0.$$

Let W' be the width obtained in two dimensions. Using the approximation of the height of the fault ellipsoid given in (2), the pdf corresponding to Expression (7) is found to be:

$$f(w|1) = \frac{\tau \rho^{1+1/\tau}}{\Gamma(1 + 1/\tau)} w^\tau \exp(-\rho w^\tau),$$

while $f(w'|w,1)$ is the same as in the fractal case, Expression (55). The integral in Expression (8) gives the pdf of the width in two dimensions:

$$f(w'|1) = \frac{\tau\rho^{1+1/\tau}}{\Gamma(1+1/\tau)} w' \int_{w'}^{\infty} \frac{w^{\tau-1}}{\sqrt{w^2-(w')^2}} \exp(-\rho w^{\tau})\, dw, \tag{57}$$

which is the pdf given in (29). For $\tau = 1$ the integral becomes

$$f(w'|1) = \int_{w'}^{\infty} \frac{w'}{w\sqrt{w^2-(w')^2}} \rho^2 w \exp(-\rho w)\, dw = \rho^2 w' \int_{w'}^{\infty} \frac{1}{\sqrt{w^2-(w')^2}} \exp(-\rho w)\, dw.$$

Using the substitution $w = w'z$ the integral can be written as

$$f(w'|1) = \rho^2 w' \int_{1}^{\infty} \frac{1}{\sqrt{z^2-1}} \exp(-\rho w' z)\, dz = \rho^2 w' K_0(\rho w'),$$

where $K_0(\cdot)$ is the Bessel function described in Appendix C. Substituting W' by W and $f(w')$ by $f_2(w)$ and omitting conditioning on $I = 1$, Expression (30) is obtained.

C Special Functions

The gamma function, incomplete gamma function and the modified Bessel function are described below, see for example Abramowitz and Stegun (1965).

C.1 Gamma function and incomplete gamma function

The gamma function $\Gamma(\cdot)$ and the incomplete gamma function $\Gamma(\cdot,\cdot)$ is given by the integrals

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t}\, dt, \qquad \Gamma(n,a) = \int_a^{\infty} t^{n-1} e^{-t}\, dt.$$

C.2 Modified Bessel function

The modified Bessel function of order ν, $K_\nu(\cdot)$, is the solution of the differential equation

$$t^2 \frac{d^2 w}{dt^2} + t\frac{dw}{dt} - (t^2+\nu^2)w = 0,$$

and can be expressed by the integral

$$K_\nu(t) = \frac{\sqrt{\pi}(t/2)^\nu}{\Gamma(\nu+1/2)} \int_1^{\infty} (z^2-1)^{\nu-1/2} e^{-tz}\, dz.$$

For $\nu = 0, 1$ the expression simplifies to

$$\begin{aligned} K_0(t) &= \int_1^\infty \frac{1}{\sqrt{z^2-1}} \cdot e^{-tz}\, dz, \\ K_1(t) &= t \int_1^\infty \sqrt{z^2-1} \cdot e^{-tz}\, dz. \end{aligned}$$

For large $|t|$, the modified Bessel function of order 0 can be approximated as

$$K_0(t) \approx \sqrt{\frac{\pi}{2t}} e^{-t} \left(1 + \frac{(4-1)}{8t} + \frac{(4-1)(4-9)}{2!(8t)^2} + \frac{(4-1)(4-9)(4-25)}{3!(8t)^3} + \cdots \right) \quad (58)$$

see Abramowitz and Stegun (1965).

Part II

Sampling algorithm for estimating Bayes factor

Sampling algorithm for estimating Bayes factor

Hilde G. Borgos

Abstract

Bayes factor can be used to choose between two specified models, based on available observations, without any requirement of nested models. In the case of equal prior probabilities of the two models, Bayes factor is given as the posterior odds ratio of the models. Evaluation of Bayes factor includes solving of integrals, which in most cases can only be solved numerically. This work presents an McMC algorithm constructed to estimate Bayes factor when the available observations contain sampling errors. The McMC algorithm allows simultaneous sampling of the model parameters. The posterior distribution of the unknown, high-dimensional, underlying variables is assumed to be of a form where the bulk of the variables are analytically tractable, while only a low-dimensional subset needs to be sampled from by the use of McMC. The performance of the algorithm is studied by considering a special case of exact observations where Bayes factor is analytically tractable, thus the estimates can be compared to the corresponding analytical values.

1 Introduction

In the work presented by Borgos (2000), two possible distributions of the size of geological faults are studied, where the size is represented as either maximum displacement or horizontal extent of the fault. It is of interest to decide which distribution is most likely to be the correct one, based on available observations. Observations of faults are often non-exact, and knowledge of the sampling procedure should be incorporated when defining the model choice problem. In addition to the model choice, it is of interest to make inference about the underlying fault population. This inference is based on the posterior distribution of the model parameters and the fault sizes.

Bayes factor can be used as a criterion for model choice, see Kass and Raftery (1995) for an overview. In many problems encountered, Bayes factor can not be found analytically, but must be approximated using numerical integration techniques, see Evans and Swartz (1995). Among the techniques available for numerical approximations, Markov chain Monte Carlo (McMC) sampling is well suited for dealing with complex, high-dimensional distributions, see for instance Besag et al. (1995), Han and Carlin (2000). McMC techniques

can be used to estimate Bayes factor, for example by including a model indicator in the sampling, as described by Carlin and Chib (1995).

The current paper presents an McMC algorithm constructed with the aim of solving the model choice problem encountered by Borgos (2000). The unknown quantities involved in this work consist of a stochastic variable, and model parameters describing the distribution of the variable under each model. The dimension of the unknown stochastic variable is high, and possibly unknown. The dimensions of the model parameters are low under both competing models, and in general need not coincide. The posterior distribution of the high-dimensional variable is assumed to be analytically tractable, and under both competing models inference about the model parameters can be based on posterior distributions obtained by an analytical integration over the high-dimensional variable. Thus the high-dimensional variable need not be included in the sample space of the McMC algorithm. General model choice problems of this kind are considered.

A Metropolis-Hastings algorithm is presented, where the target distribution is defined inspired by the target distribution of the Gibbs algorithm presented by Carlin and Chib (1995), including a model indicator as a parameter. The target distribution is constructed with the aim of estimating Bayes factor and the model parameters of the preferred model. The target distribution is not uniquely defined however, but leaves some freedom of choice that will affect neither the value of Bayes factor nor the posterior distribution of the model parameter under the preferred model. The freedom of choice in the definition of the target distribution is exploited to optimize the performance of the McMC algorithm. The estimator of Bayes factor is expressed through a statistic of the realizations from the McMC algorithm. Minimization of uncertainty of the estimator of Bayes factor is complicated, but instead minimization of the uncertainty of the statistic involved in the estimator is used as a criterion for optimality.

The general model choice problem is presented in Section 2, and Section 3 gives a discussion on estimation of Bayes factor based on McMC sampling. The model choice problem encountered by Borgos (2000) is summarized in Section 4, and in Section 5 the general behavior of Bayes factor and the performance of the McMC algorithm on this particular problem is evaluated.

2 Model Choice Problem

A high-dimensional stochastic variable $\boldsymbol{Y}$ is believed to originate from one of two possible models, where the stochastic variable $M \in \Omega_m = \{m_1, m_2\} = \{1, 2\}$ indicates which model is the true one, model 1 or 2 respectively. Under model m, the probability density function (pdf) of $\boldsymbol{Y}$ is defined as $f(\boldsymbol{y}|m, \psi_m)$, where $\Psi_m \in \Omega_{\psi_m}$ is a low-dimensional stochastic vector of model parameters under model m. The dimensions of the parameters Ψ_1 and Ψ_2

under the two models need not coincide. The model parameter of model m is assumed to have a proper prior $f(\psi_m|m)$, thus the joint distribution of $\boldsymbol{Y}$ and Ψ_m is

$$f(\boldsymbol{y}, \psi_m|m) = f(\boldsymbol{y}|m, \psi_m)f(\psi_m|m). \tag{1}$$

The aim of this work is to decide which model is most likely, and to make inference about the model parameter Ψ_m corresponding to the favored model m and the variable $\boldsymbol{Y}$.

The parameter Ψ_m is unknown and the variables $\boldsymbol{Y}$ are not observed exactly, but inference can be made based on observations $\boldsymbol{X}$. The observations can include different types of errors, like precision errors, bias and censoring. After a model choice has been made, the posterior pdfs of interest are the conditional pdfs $f(\boldsymbol{y}, \psi_m|\boldsymbol{x}, m) = f(\boldsymbol{y}|\boldsymbol{x}, m, \psi_m)f(\psi_m|\boldsymbol{x}, m)$, where m is the preferred model. The observations $\boldsymbol{X}$ are described by a likelihood function $h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m)$ under model m, depending on both the high-dimensional variable $\boldsymbol{Y}$ and the low-dimensional model parameter Ψ_m. However, the likelihood function and prior distributions are assumed to be of a form where

$$h(\boldsymbol{x}|m, \psi_m) = \int h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m)f(\boldsymbol{y}|m, \psi_m)\, d\boldsymbol{y} \tag{2}$$

is analytically tractable. Thus the likelihood function can be reduced to only depending on a small set of variables. The posterior distribution of Ψ_m under model m can now be expressed as

$$f(\psi_m|\boldsymbol{x}, m) = \text{const} \times h(\boldsymbol{x}|m, \psi_m)f(\psi_m|m), \tag{3}$$

and inference about Ψ_m can be made based on an analytical randomization over $\boldsymbol{Y}$. The posterior distribution of $\boldsymbol{Y}$ under model m, conditioned on Ψ_m, is given by

$$f(\boldsymbol{y}|\boldsymbol{x}, m, \psi_m) = \frac{h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m)f(\boldsymbol{y}|m, \psi_m)}{h(\boldsymbol{x}|m, \psi_m)}. \tag{4}$$

Bayes factor can be used as a criterion for model choice when two or more alternative models are suggested, without any requirement of nested models. See Kass and Raftery (1995) for an overview of Bayes factor. In the case of two competing models, the two outcomes $m_1 = 1$ and $m_2 = 2$ of the indicator variable M are assigned prior probabilities $f(m_1)$ and $f(m_2)$, which are assumed to be known. Bayes factor is defined as the ratio between the posterior odds ratio and the prior odds ratio:

$$B_{12}(\boldsymbol{x}) = \frac{f(m_1|\boldsymbol{x})}{f(m_2|\boldsymbol{x})} \cdot \frac{f(m_2)}{f(m_1)} = \frac{f(\boldsymbol{x}|m_1)}{f(\boldsymbol{x}|m_2)}. \tag{5}$$

If there is no prior reason to prefer one model to the other, the prior distributions are chosen as $f(m_1) = f(m_2)$ and Bayes factor equals the posterior odds ratio. If $B_{12}(\boldsymbol{x}) > 1$ the observations favor model 1, while if $B_{12}(\boldsymbol{x}) < 1$ the observations support model 2. The

marginal density function $f(\boldsymbol{x}|m)$ under model m is defined as

$$\begin{aligned} f(\boldsymbol{x}|m) &= \iint h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m) f(\boldsymbol{y}|m, \psi_m) f(\psi_m|m)\, d\boldsymbol{y}\, d\psi_m \\ &= \int h(\boldsymbol{x}|m, \psi_m) f(\psi_m|m)\, d\psi_m. \end{aligned} \tag{6}$$

From Expressions (5) and (6) it is observed that the value of Bayes factor does not depend on the choice of prior probabilities $f(m)$. Bayes factor can be interpreted as the ratio between two normalizing constants, since $f(\boldsymbol{x}|m)$ is given as the normalizing constant in the relationship in Expression (3):

$$f(\psi_m|\boldsymbol{x}, m) = \frac{h(\boldsymbol{x}|m, \psi_m) f(\psi_m|m)}{f(\boldsymbol{x}|m)}. \tag{7}$$

The marginal distribution in Expression (6) is analytically tractable only in specific cases. In general, Bayes factor can not be analytically obtained, but must instead be estimated. Estimation of Bayes factor is the main topic of this report.

The value of Bayes factor may depend strongly on the values of the hyper-parameters in the prior distributions of Ψ_1 and Ψ_2. As discussed by Kass and Raftery (1995), the choice of hyper-parameters can be critical to the result of the model choice. Thus strong prior information about the parameters are desirable to give a reliable conclusion of the model choice. Furthermore, only proper priors should be used. If improper priors are specified for the model parameters, Bayes factor is only found up to an unknown constant. Alternative analytical definitions of Bayes factors are suggested, giving factors less sensitive to the choice of prior distribution. Some examples are the posterior Bayes factor (Aitkin, 1991), partial Bayes factor and fractional Bayes factor (O'Hagan, 1991, 1995) and the intrinsic Bayes factor (Berger and Pericchi, 1996). A common feature of these alternative Bayes factors is that they combine an initial prior, which may be improper, with some or all of the observations to obtain a proper prior for the parameter Ψ_m. In this work the priors $f(\psi_m|m)$ are assumed to be well specified proper priors, and the original definition of Bayes factor in Expression (5) is used.

3 Estimating Bayes Factor

The integrals in Expression (6), involved in the expression of Bayes factor, are in general not analytically tractable. If no analytical solution is found, Bayes factor can be calculated by numerical integration, using for example asymptotic approximations or sampling approaches like simple Monte Carlo, importance sampling or McMC sampling. Evans and Swartz (1995) and Kass and Raftery (1995) discuss different estimation techniques. In this work the focus is on estimation techniques based on McMC sampling. See, for example,

Besag et al. (1995) and references therein for an introduction to McMC theory. Han and Carlin (2000) gives an overview of McMC algorithms constructed to estimate Bayes factor.

The aim of the current work is to calculate Bayes factor in order to choose between two proposed models, and to make inference about the model parameter Ψ_m and the unknown variable $\boldsymbol{Y}$ under the favored model m, see Expressions (3) and (4). McMC techniques enable simultaneous sampling of both the model indicator M and the model parameters Ψ_1 and Ψ_2. The variable $\boldsymbol{Y}$ is not included in the McMC algorithm, but samples of the variable $\boldsymbol{Y}$ can be generated from the posterior distribution in Expression (4) after the model choice is performed and the corresponding model parameter is estimated.

Realizations from the posterior distribution of one specific model can be used to estimate Bayes factor, see Kass and Raftery (1995) for an overview. Chib (1995) and DiCiccio et al. (1997) use the fact that Bayes factor is given as a ratio between normalizing constants in posterior distributions, see Expressions (5) and (7). The normalizing constant under the specified model can be estimated based on the realizations, and by doing this for both models under consideration an estimate of Bayes factor can be obtained. Other algorithms constructed to estimate normalizing constants or ratios between normalizing constants can also be used to estimate Bayes factor, see Gelman and Meng (1998) for an overview.

Instead of considering one model at a time, the model indicator M can be included as a variable in the McMC algorithm. The Reversible Jump McMC algorithm suggested by Green (1995) generate realizations of (M, Ψ_M) from $f(m, \psi_m|\boldsymbol{x})$. The sample space of the algorithm is defined as $\Omega_m \times (\Omega_{\psi_1} \cup \Omega_{\psi_2})$. Carlin and Chib (1995) described a Gibbs sampler on the sample space $\Omega_m \times \Omega_{\psi_1} \times \Omega_{\psi_2}$, giving realizations of (M, Ψ_1, Ψ_2) from the posterior pdf $f(m, \psi_1, \psi_2|\boldsymbol{x})$. In order for this pdf to be completely defined, pdfs $\pi(\psi_{m_*}|m)$, $m_* \neq m$, must be specified. Carlin and Chib (1995) denote these pdfs pseudopriors or linking densities, which are defined under the assumption that Ψ_{m_*} is independent of Ψ_m and $\boldsymbol{X}$, given $M = m$. Dellaportas et al. (1998) suggested combining the two above mentioned algorithms. In the Gibbs algorithm of Carlin and Chib (1995), update steps of the model indicator are replaced by an acceptance/rejection step based on the Reversible Jump algorithm of Green (1995). Other sampling algorithms based on the idea of including M as a variable are presented by Carlin and Polson (1991) and George and McCulloch (1993).

3.1 McMC algorithm

In this work Bayes factor is estimated by the use of an McMC algorithm where the model indicator M is included in the sampling scheme. A target distribution $\pi(m, \psi_m, \psi_{m_*}|\boldsymbol{x})$, $m_* \neq m$, is defined, adopting the concept of pseudopriors. This pdf is defined with the aim of estimating Bayes factor and the model parameter of the favored model, and it is shown below that all pseudopriors can be chosen arbitrary without altering the quantities

of interest. The target distribution is defined as

$$\begin{aligned}\pi(m,\psi_m,\psi_{m_*}|\boldsymbol{x}) &= \pi(\boldsymbol{x})^{-1}h(\boldsymbol{x}|m,\psi_m,\psi_{m_*})f(\psi_m|m)\pi(\psi_{m_*}|m)\pi(m)\\ &= \pi(\boldsymbol{x})^{-1}h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\pi(\psi_{m_*}|m)\pi(m). \qquad (8)\end{aligned}$$

Throughout this work the notation $\pi(\cdot)$ is used do denote pseudopriors or to stress that a pdf is defined with respect to the pseudopriors. In the likelihood function $h(\boldsymbol{x}|m,\psi_m,\psi_{m_*})$, M is an indicator of which Ψ_m applies to the observations, and the likelihood function is assumed to be independent of Ψ_{m_*}. Furthermore, Ψ_m and Ψ_{m_*} are assumed to be independent given M. Since Bayes factor is independent of the actual value of the prior pdf $f(m)$, see Expression (5), any proper discrete pdf with $\pi(m) > 0$ for all $m \in \Omega_m$ can be used as a pseudoprior for M in the target distribution (8). As pointed out by Carlin and Chib (1995), the pseudoprior $\pi(\psi_{m_*}|m)$, $m_* \neq m$, can be chosen as any proper pdf with $\pi(\psi_{m_*}|m) > 0$ for all $\psi_{m_*} \in \Omega_{\psi_{m_*}}$. This is illustrated through the expression

$$\begin{aligned}\pi(\boldsymbol{x}|m) &= \iint \pi(m)^{-1}\pi(m,\psi_m,\psi_{m_*}|\boldsymbol{x})\pi(\boldsymbol{x})\,d\psi_m\,d\psi_{m_*}\\ &= \iint h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\pi(\psi_{m_*}|m)\,d\psi_m\,d\psi_{m_*}\\ &= \int h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\,d\psi_m,\\ &= f(\boldsymbol{x}|m) \qquad (9)\end{aligned}$$

showing that $\pi(\boldsymbol{x}|m_1)/\pi(\boldsymbol{x}|m_2)$ gives the correct value of Bayes factor regardless of the form of the pseudopriors, see Expression (5). Furthermore, the posterior pdf

$$\begin{aligned}\pi(\psi_m|\boldsymbol{x},m) &= \int \pi(m|\boldsymbol{x})^{-1}\pi(m,\psi_m,\psi_{m_*}|\boldsymbol{x})\,d\psi_{m_*}\\ &= \int (\pi(m|\boldsymbol{x})\pi(\boldsymbol{x}))^{-1}h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\pi(\psi_{m_*}|m)\pi(m)\,d\psi_{m_*}\\ &= \int \pi(\boldsymbol{x}|m)^{-1}h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\pi(\psi_{m_*}|m)\,d\psi_{m_*}\\ &= f(\boldsymbol{x}|m)^{-1}\,h(\boldsymbol{x}|m,\psi_m)f(\psi_m|m)\\ &= f(\psi_m|\boldsymbol{x},m) \qquad (10)\end{aligned}$$

is also correct, using Expression (9), and inference about Ψ_m is not affected by the pseudopriors. Finally, the posterior distribution of $\boldsymbol{Y}$ under model m, Expression (4), is also independent of the choice of pseudopriors.

Although all marginal distributions involved in the expression of Bayes factor and the posterior distributions of interest are independent of the pseudopriors, some of the other marginal pdfs may depend on them. Some examples are:

$$\begin{aligned}\pi(m|\boldsymbol{x}) &= \pi(\boldsymbol{x})^{-1}\pi(\boldsymbol{x}|m)\pi(m) \;=\; \pi(\boldsymbol{x})^{-1}f(\boldsymbol{x}|m)\pi(m)\\ \pi(\psi_m|\boldsymbol{x}) &= f(\psi_m|\boldsymbol{x},m)\pi(m|\boldsymbol{x}) + \pi(\psi_m|m_*)\pi(m_*|\boldsymbol{x}).\end{aligned}$$

The introduction of pseudopriors $\pi(\psi_m|m_*)$ and $\pi(\psi_{m_*}|m)$ on the model parameters is required for the posterior distribution of (M, Ψ_m, Ψ_{m_*}) to be completely specified. Furthermore, Expression (5) implies the possibility of introducing a pseudoprior $\pi(m)$ also for the model indicator, without altering the results. These pseudopriors constitute a freedom of choice in the specification of the target distribution (8), which will be exploited to minimize the uncertainty in the estimate of Bayes factor. The uncertainty of the estimate depends on the behavior of the McMC algorithm. Before optimal choices of pseudopriors can be discussed, the McMC algorithm is presented and an estimator of Bayes factor is suggested.

A Metropolis-Hastings algorithm is constructed to sample from the target distribution in Expression (8), see Hastings (1970). The algorithm involves sampling of the model indicator M and the model parameters Ψ_1 and Ψ_2, conditioned on observations $\boldsymbol{X}$. Three transition steps are defined within one iteration of the algorithm, one for each variable. The McMC algorithm for estimating Bayes factor based on observations $\boldsymbol{X}$ can be expressed as follows:

Algorithm 1

Let $\Omega = \Omega_m \times \Omega_{\psi_m} \times \Omega_{\psi_{m_*}}$ be the sample space of (M, Ψ_m, Ψ_{m_*}) and assume the target distribution in Expression (8) satisfies $\pi(m, \psi_m, \psi_{m_*}|\boldsymbol{x}) > 0; \forall (m, \psi_m, \psi_{m_*}) \in \Omega$.

- *Initiate*
  - *Arbitrary $(m(0), \psi_m(0), \psi_{m_*}(0)) \in \Omega$.*
- *Iterate, $t = 1, 2, \ldots$*
  - Step 1*:*
    - *Denote $m(t-1) = m$, $\psi_m(t-1) = \psi_m$ and $\psi_{m_*}(t-1) = \psi_{m_*}$.*
    - *Propose m' from $q(m'|m)$.*
    - *Calculate*
      for $m' = m_* \neq m$:

$$\alpha(m_*|m) = \min\left\{1, \frac{\pi(m_*)q(m|m_*)}{\pi(m)q(m_*|m)} \times \frac{h(\boldsymbol{x}|m_*, \psi_{m_*})f(\psi_{m_*}|m_*)}{\pi(\psi_{m_*}|m)} \times \frac{\pi(\psi_m|m_*)}{h(\boldsymbol{x}|m, \psi_m)f(\psi_m|m)}\right\}, \tag{11}$$

      for $m' = m$: $\alpha(m|m) = 1$
    - *Let: $m(t) = m'$ with probability $\alpha(m'|m)$,*
      $m(t) = m$ else.

– Step 2*:*

  * *Denote* $m(t) = m$ *and* $\psi_m(t-1) = \psi_m$.
  * *Propose* ψ'_m *from* $q(\psi'_m|\psi_m, m)$.
  * *Calculate*

$$\alpha(\psi'_m|\psi_m, m) \;=\; \min\left\{1, \frac{h(\boldsymbol{x}|m, \psi'_m) f(\psi'_m|m) q(\psi_m|\psi'_m, m)}{h(\boldsymbol{x}|m, \psi_m) f(\psi_m|m) q(\psi'_m|\psi_m, m)}\right\}. \tag{12}$$

  * *Let:* $\psi_m(t) = \psi'_m$ *with probability* $\alpha(\psi'_m|\psi_m, m)$,
    $\psi_m(t) = \psi_m$ *else.*

– Step 3*:*

  * *Denote* $m(t) = m$ *and* $\psi_{m_*}(t-1) = \psi_{m_*}$ *where* $m_* \neq m$.
  * *Propose* ψ'_{m_*} *from* $q(\psi'_{m_*}|\psi_{m_*}, m)$.
  * *Calculate*

$$\alpha(\psi'_{m_*}|\psi_{m_*}, m) \;=\; \min\left\{1, \frac{\pi(\psi'_{m_*}|m) q(\psi_{m_*}|\psi'_{m_*}, m)}{\pi(\psi_{m_*}|m) q(\psi'_{m_*}|\psi_{m_*}, m)}\right\}. \tag{13}$$

  * *Let:* $\psi_{m_*}(t) = \psi'_{m_*}$ *with probability* $\alpha(\psi'_{m_*}|\psi_{m_*}, m)$,
    $\psi_{m_*}(t) = \psi_{m_*}$ *else.*

The variables $(M(t), \Psi_m(t), \Psi_{m_*}(t))$ at iteration t of the algorithm have marginal pdf $\pi^t(m, \psi_m, \psi_{m_*}|\boldsymbol{x})$ where $\lim_{t\to\infty} \pi^t(m, \psi_m, \psi_{m_*}|\boldsymbol{x}) = \pi(m, \psi_m, \psi_{m_*}|\boldsymbol{x})$. The systematic scan of the three steps of the algorithm does not imply a time-reversible Markov chain, but the chain is both aperiodic and irreducible. Lack of time-reversibility has no effect on the estimates presented below. Time-reversibility can be obtained by drawing at random the order of the steps at each iteration. Note that if the pseudoprior is used as transition density for Ψ_{m_*}, $q(\psi'_{m_*}|\psi_{m_*}, m) = \pi(\psi'_{m_*}|m)$, the acceptance probability (13) becomes $\alpha(\psi'_{m_*}|\psi_{m_*}, m) = 1$ and the change is always accepted.

Both Bayes factor $B_{12}(\boldsymbol{x})$ and the model parameters Ψ_1 and Ψ_2 can be estimated based on realizations obtained from McMC simulations. Suppose that n subsequent samples $\{m^j, \psi_1^j, \psi_2^j; j = 1, \ldots, n\}$ are gathered after convergence of the algorithm is reached. Estimation of $B_{12}(\boldsymbol{x})$ is essentially the same as estimating the posterior pdf $\pi(m_1|\boldsymbol{x})$, and a natural estimator of the latter is given by the relative number of outcomes of model 1:

$$\hat{\pi}(m_1|\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[M^j = m_1]. \tag{14}$$

The marginal distribution of M^j is $\pi(m|\boldsymbol{x})$, thus the mean and variance of the estimator

are given as

$$\begin{aligned}
\mathrm{E}\left[\hat{\pi}(m_1|\boldsymbol{x})\right] &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left[\mathbf{1}[M^j = m_1]\right] = \pi(m_1|\boldsymbol{x}) \\
\mathrm{Var}\left[\hat{\pi}(m_1|\boldsymbol{x})\right] &= \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}\left[\mathbf{1}[M^j = m_1]\right] + \frac{2}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n-j}\mathrm{Cov}[\mathbf{1}[M^j = m_1], \mathbf{1}[M^{j+k} = m_1]] \\
&= \frac{1}{n}\pi(m_1|\boldsymbol{x})(1-\pi(m_1|\boldsymbol{x})) + \frac{2}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n-j}\gamma(k), \qquad (15)
\end{aligned}$$

where $\gamma(k) = \mathrm{Cov}[\mathbf{1}[M^j = m_1], \mathbf{1}[M^{j+k} = m_1]]$ is the autocovariance at lag k. A natural estimator of $B_{12}(\boldsymbol{x})$, see Expression (5), is given as

$$\hat{B}_{12}(\boldsymbol{x}) = \frac{\hat{\pi}(m_1|\boldsymbol{x})}{1-\hat{\pi}(m_1|\boldsymbol{x})} \cdot \frac{\pi(m_2)}{\pi(m_1)}. \qquad (16)$$

Recall that from the definition of $B_{12}(\boldsymbol{x})$, its true value is independent of the pseudoprior $\pi(m)$. However, the distribution of the estimator $\hat{B}_{12}(\boldsymbol{x})$ depends on the choice of $\pi(m)$. Note that the mean and variance of $\hat{B}_{12}(\boldsymbol{x})$

$$\begin{aligned}
\mathrm{E}\left[\hat{B}_{12}(\boldsymbol{x})\right] &= \frac{\pi(m_2)}{\pi(m_1)}\mathrm{E}\left[\frac{\hat{\pi}(m_1|\boldsymbol{x})}{1-\hat{\pi}(m_1|\boldsymbol{x})}\right] \\
\mathrm{Var}\left[\hat{B}_{12}(\boldsymbol{x})\right] &= \left(\frac{\pi(m_2)}{\pi(m_1)}\right)^2 \mathrm{Var}\left[\frac{\hat{\pi}(m_1|\boldsymbol{x})}{1-\hat{\pi}(m_1|\boldsymbol{x})}\right] \qquad (17)
\end{aligned}$$

need not be finite, since $\hat{\pi}(m_1|\boldsymbol{x})$ may equal 1.

An estimator of the parameter value in model m is given by the empirical posterior mean

$$\hat{\Psi}_m = \hat{\mathrm{E}}\left[\Psi_m|\boldsymbol{x}, m\right] = \frac{\sum_{j=1}^{n}\Psi_m^j \mathbf{1}[M^j = m]}{\sum_{j=1}^{n}\mathbf{1}[M^j = m]}. \qquad (18)$$

The subset $\{\psi_m^j | m^j = m\}$ of realizations from the McMC algorithm can also be used to describe the complete posterior pdf $f(\psi_m|\boldsymbol{x}, m)$. Finally, realizations of $\boldsymbol{Y}$ under the favored model m can be obtained by sampling from the conditional posterior distribution $f(\boldsymbol{y}|\boldsymbol{x}, m, \psi_m)$ given in Expression (4). An example is given in Section 4.2, where it is explained how $f(\boldsymbol{y}|\boldsymbol{x}, m, \psi_m)$ can be sampled from by the use of rejection sampling.

3.2 Pseudopriors

The pseudopriors $\pi(\psi_1|m_2)$, $\pi(\psi_2|m_1)$ and $\pi(m)$ can be chosen as any proper pdfs satisfying $\pi(\psi_m|m_*) > 0; \forall \psi_m \in \Omega_{\psi_m}$, $m = m_1, m_2$, and $\pi(m) > 0; \forall m \in \Omega_m$. In search for good

pseudopriors, the major goal is to reduce uncertainty in the estimator of Bayes factor. The pseudopriors are chosen focusing on the behavior of M in the McMC algorithm. In addition, the transition densities $q(\psi'_m|\psi_m, m)$ and $q(\psi'_{m_*}|\psi_{m_*}, m)$ in Algorithm 1 should be chosen to span the sample space of the model parameters Ψ_1 and Ψ_2 adequately.

The estimator (16) of Bayes factor need not have a finite mean and variance, see Expression (17), and this complicates evaluation of the properties of the estimator. The precision of $\hat{B}_{12}(\boldsymbol{x})$ is influenced by the properties of the estimator $\hat{\pi}(m_1|\boldsymbol{x})$, see Expression (14), and minimization of the uncertainty of the latter estimator is considered. If the variables M^j are independent, $\gamma(k) = 0$ for $k \neq 0$ in Expression (15), and $\mathrm{Var}\left[\hat{\pi}(m_1|\boldsymbol{x})\right] = n^{-1}\pi(m_1|\boldsymbol{x})(1 - \pi(m_1|\boldsymbol{x}))$. However, the variance of of $\hat{\pi}(m_1|\boldsymbol{x})$ is further reduced if the autocovariance at lag 1 is negative, $\gamma(1) < 0$, giving a negative correlation between succeeding realizations of M. The extreme case is $\gamma(1) = -\mathrm{Var}\left[\mathbf{1}[M = m_1]\right] = -\pi(m_1|\boldsymbol{x})(1 - \pi(m_1|\boldsymbol{x}))$, where M alternates periodically between m_1 and m_2, a situation that arises if the probability of a move from m to m' in one step of Algorithm 1 satisfies:

$$\mathrm{P}\{M(t) = m'|M(t-1) = m\} = q(m'|m)\alpha(m'|m) = \begin{cases} 0, & m' = m \\ 1, & m' = m_* \neq m. \end{cases} \tag{19}$$

If this extreme case should actually be obtained, the periodicity of the Markov chain can be removed through minor adjustments of $q(m'|m)$. However, in practice the requirement will never be exactly reached, and periodicity is no problem. Although probably never obtained, this extreme case still serves as a target for optimal achievement. The requirement (19) can only be fulfilled if

$$\alpha(m_*|m) = 1, \quad q(m'|m) = \mathbf{1}[m' \neq m] = \begin{cases} 0, & m' = m \\ 1, & m' = m_* \neq m. \end{cases}$$

Thus the acceptance probability in Expression (11) should satisfy

$$\alpha(m_*|m) = \min\left\{1, \frac{\pi(m_*)}{\pi(m)} \cdot \frac{h(\boldsymbol{x}|m_*, \psi_{m_*}) f(\psi_{m_*}|m_*)}{\pi(\psi_{m_*}|m)} \cdot \frac{\pi(\psi_m|m_*)}{h(\boldsymbol{x}|m, \psi_m) f(\psi_m|m)}\right\} = 1, \tag{20}$$

while $\alpha(m|m) = 1$ always holds in Algorithm 1. The two last ratios in $\alpha(m_*|m)$ are independent of the model parameters if their pseudopriors are chosen equal to the posterior pdfs,

$$\pi(\psi_1|m_2) = f(\psi_1|\boldsymbol{x}, m_1), \qquad \pi(\psi_2|m_1) = f(\psi_2|\boldsymbol{x}, m_2). \tag{21}$$

The two ratios are then given as

$$\frac{h(\boldsymbol{x}|m_*, \psi_{m_*}) f(\psi_{m_*}|m_*)}{f(\psi_{m_*}|\boldsymbol{x}, m_*)} \cdot \frac{f(\psi_m|\boldsymbol{x}, m)}{h(\boldsymbol{x}|m, \psi_m) f(\psi_m|m)} = \frac{f(\boldsymbol{x}|m_*)}{f(\boldsymbol{x}|m)} = B_{m_*m}(\boldsymbol{x}), \tag{22}$$

resulting in the requirement

$$\alpha(m_*|m) = \min\left\{1, \frac{\pi(m_*)}{\pi(m)} B_{m_*m}(\boldsymbol{x})\right\} = 1, \tag{23}$$

which is fulfilled if the pseudoprior of M is chosen as

$$\pi(m_1) = (1 + B_{12}(\boldsymbol{x}))^{-1}. \tag{24}$$

Thus the requirement in Expression (19) is obtained if the transition density is chosen as $q(m'|m) = \mathbf{1}[m' \neq m]$ and the pseudopriors are chosen as given in Expressions (21) and (24). In this case the variance of the estimator $\hat{\pi}(m_1|\boldsymbol{x})$ is minimized. The fulfillment of the requirement in Expression (23) implies $\pi(m|\boldsymbol{x}) = \pi(m_*|\boldsymbol{x}) = 0.5$, which coincides with the estimate $\hat{\pi}(m|\boldsymbol{x}) = 0.5$ obtained when M alternates periodically between m and m_*.

The optimal choice of pseudopriors is expressed through Bayes factor and the posterior pdfs of the model parameters, which are the unknown quantities the algorithm is constructed to estimate. Thus an optimal algorithm can only be constructed if the answers are known beforehand, in which case there would be no need for the McMC sampling. However, the knowledge of the optimal choice can serve as a guideline for how to choose the pseudopriors, by choosing pdfs sufficiently close to the optimal solution. One way of doing this is to run an McMC algorithm sampling independently from $f(\psi_1|\boldsymbol{x}, m_1)$ and $f(\psi_2|\boldsymbol{x}, m_2)$, and use the realizations of Ψ_1 and Ψ_2 to estimate normalized posterior pdfs and to give a rough estimate of Bayes factor. This sampling can be performed by Algorithm 1, using only the transition kernel $q(\psi'_m|\psi_m, m)$. If r realizations Ψ_m^j, $j = 1, \ldots, r$ are generated, the pseudoprior of Ψ_m can be chosen as a kernel density estimator

$$\pi(\psi_m|m_*) = \hat{f}(\psi_m|\boldsymbol{x}, m) = \frac{1}{r}\sum_{j=1}^{r} g(\psi_m|\psi_m^j, h_m), \tag{25}$$

where the kernel $g(\psi_m|\psi_m^j, h_m)$ is a proper pdf parameterized by ψ_m^j and h_m, see Silverman (1986). By including the observations ψ_m^j in the parameterization, the kernel densities can be scaled to give normalized densities on Ω_{ψ_m}, and the pseudoprior is defined as a proper pdf conditioned on $\Psi_m \in \Omega_{\psi_m}$. The pseudoprior of M in Expression (24) is based on a preliminary estimate of Bayes factor, given by

$$\tilde{B}_{12}(\boldsymbol{x}) = \frac{1}{L}\sum_{j=1}^{L} \frac{h(\boldsymbol{x}|m_1, \tilde{\psi}_1^j) f(\tilde{\psi}_1^j|m_1)}{\pi(\tilde{\psi}_1^j|m_2)} \cdot \frac{\pi(\tilde{\psi}_2^j|m_1)}{h(\boldsymbol{x}|m_2, \tilde{\psi}_2^j) f(\tilde{\psi}_2^j|m_2)}, \tag{26}$$

where $\tilde{\psi}_m^j$ are arbitrary values of Ψ_m, picked from the set of realizations. This is a similar approach as used by Chib (1995), who emphasize that normalizing constants can be estimated by evaluating the likelihood function, prior pdf and estimated posterior pdf only for one parameter value. From Expression (22) it is observed that if the optimal choices of pseudopriors $\pi(\psi_m|m_*)$ are used, Expression (21), the estimate $\tilde{B}_{12}(\boldsymbol{x})$ will give an exact solution regardless of the values of $\tilde{\psi}_m^j$. If the pseudopriors $\pi(\psi_m|m_*)$ are sufficiently close to the posterior pdfs, the rough estimate (26) can be based on a single value $\tilde{\psi}_m$, i.e., $L = 1$. In this case, when the pseudopriors from Expression (25) are used the estimate of the posterior pdf is assumed to be best near the mode of $\pi(\psi_m|m_*)$, and $\tilde{\psi}_m$ should be chosen from this area.

The choice of pseudopriors aims at obtaining negatively correlated realizations of M, to minimize the uncertainty in the estimator $\hat{\pi}(m_1|\boldsymbol{x})$. Since the optimal choice of pseudopriors can not be reached, the extreme case with $q(m'|m) = \mathbf{1}[m' \neq m]$ and $\alpha(m'|m) = 1$ is not fulfilled. The transition density $q(m'|m)$ is left unchanged, but the acceptance probability for a change of model will in practice depend on both the model M and the parameter values ψ_m and ψ_{m_*}. Note that the negative correlation obtained when using optimal pseudopriors only applies to M, and not the model parameters Ψ_m and Ψ_{m_*}. The correlation structure for the model parameters depend on the transition kernels $q(\psi'_m|\psi_m, m)$ and $q(\psi'_{m_*}|\psi_{m_*}, m)$ in Algorithm 1, which are left unspecified in the discussion above.

The McMC algorithm is run twice, first to adjust pseudopriors, next to estimate Bayes factor and the model parameter of the favored model. Note, however, that although both runs of the algorithm use the same observations, the limiting distributions of interest in the model choice problem are not influenced by this repeated use of the data. The first McMC algorithm sampling from the posterior distributions is only used to improve the performance of the algorithm by adjusting the pseudopriors and exploiting the freedom in choosing these pdfs. In the second and final run of the McMC algorithm, the limiting distributions involved in the expression of Bayes factor and the posterior pdfs of the model parameters are unaffected by the pseudopriors, see Expressions (9) and (10).

3.3 Uncertainty in estimated Bayes factor

In general the acceptance probability $\alpha(m'|m)$ is influenced by the parameter values ψ_m and ψ_{m_*}, thus it is not straightforward to quantify the uncertainty in the estimated Bayes factor. In the case of $\{M^j; j = 1, \ldots, n\}$ being independent realizations from the posterior pdf $\pi(m|\boldsymbol{x})$, an approximate confidence interval for $B_{12}(\boldsymbol{x})$ can be obtained. Then $\mathrm{P}\{M^j = m_1\} = \pi(m_1|\boldsymbol{x}) = p$ and $n\hat{\pi}(m_1|\boldsymbol{x})$ follows a binomial distribution $n\hat{\pi}(m_1|\boldsymbol{x}) \sim \mathrm{bin}(n, p)$, see Expression (14), which for large n can be approximated by a normal distribution $n\hat{\pi}(m_1|\boldsymbol{x}) \sim N(np, np(1-p))$. From Expression (5) p can be expressed as a function of $B_{12}(\boldsymbol{x})$, and combined with the normal approximation this is used to find an approximate $(1-\alpha)100\%$ confidence interval for $B_{12}(\boldsymbol{x})$:

$$\hat{B}_{12}(\boldsymbol{x}) + \frac{\pi(m_2)}{\pi(m_1)} \cdot \frac{u^2_{\alpha/2}}{2n(1-\hat{\pi}(m_1|\boldsymbol{x}))^2} \left(1 \pm \left(1 + \frac{4n\hat{\pi}(m_1|\boldsymbol{x})(1-\hat{\pi}(m_1|\boldsymbol{x}))}{u^2_{\alpha/2}} \right)^{1/2} \right), \quad (27)$$

where $u_{\alpha/2}$ is a critical value of the standard normal distribution. The interval is not centered at $\hat{B}_{12}(\boldsymbol{x})$, but it is easily shown that the estimator is covered by the interval.

If the pseudopriors are sufficiently well adapted so that negatively correlated samples of M are generated, the approximate confidence interval will be a conservative interval with too wide confidence limits. If on the other hand samples of M are positively correlated, the estimated confidence interval obtained under the assumption of independent samples is

likely to give too narrow confidence limits. In order to reduce the correlation, realizations used in the interval estimation should be separated by a number of iterations, on the expense of having fewer samples. The separation length depends on the mixing of the algorithm. Although not all samples are used to find a confidence interval for Bayes factor, the best point estimate $\hat{B}_{12}(\boldsymbol{x})$ is still obtained by using all samples.

4 Example: Geological Faults

In the work by Borgos (2000) two models are suggested for the distribution of the size of geological faults. The fault sizes of k faults are represented by a vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_k)$ of iid stochastic variables, where size is measured as maximum displacement, and under model m the pdf of $\boldsymbol{Y}$ is given by

$$f(\boldsymbol{y}|m, k, \theta_m) = \prod_{j=1}^{k} f(y_j|m, \theta_m). \tag{28}$$

The number of faults, K, is treated as a stochastic variable in cases where also the dimension of $\boldsymbol{Y}$ is unknown. A power law, or fractal, distribution is frequently used to model fault displacements, see for example JSG (1996, Vol. 18). This fractal model corresponds to the Pareto distribution,

$$f(y|m_1, \theta_1) = \frac{\theta_1 y_0^{\theta_1}}{y^{\theta_1+1}}, \qquad y \geq y_0,\ \theta_1 > 0, \tag{29}$$

which is suggested as the distribution of Y under model 1. The pareto distribution is described by two parameters, Θ_1 and y_0, where the lower limit $y_0 > 0$ must be defined in order for the pdf $f(y|m_1, \theta_1) \propto y^{-(\theta_1+1)}$ of the power law distribution to be a proper pdf. The sample space of Θ_1 is $\Omega_{\theta_1} = (0, \infty)$, while the lower limit y_0 is assumed to be known. As a competing model, a truncated exponential distribution is suggested:

$$f(y|m_2, \theta_2) = \theta_2 \exp(-\theta_2(y - y_0)), \qquad y \geq y_0,\ \theta_2 > 0. \tag{30}$$

The exponential distribution is described by the parameter Θ_2, with sample space $\Omega_{\theta_2} = (0, \infty)$. Exponential distributions of fault displacements have been observed, for example, on mid-ocean ridges, see Cowie et al. (1994), and in clay models, see Spyropoulos et al. (1999) and Ackermann et al. (1999). A common lower bound $y_0 > 0$ is used for the two competing distributions. This enables calculation of Bayes factor based on observations where the limit of resolution has imposed a lower bound on observable values.

It is of interest to estimate true fault sizes $\boldsymbol{Y}$ based on observations $\boldsymbol{X}$, including estimation of K if the dimension of $\boldsymbol{Y}$ is unknown. Two alternative sampling regimes are considered. In Section 4.1 the case of exact observations is discussed, while the more general case of non-exact observations is discussed in Section 4.2. In Section 4.3, Algorithm 1 is adapted to the model choice problem for fault size distribution.

4.1 Exact observations

The simplest sampling situation is the case of exact observations $\boldsymbol{X} \equiv \boldsymbol{Y}$, where all k variables are observed without any measurement errors. The likelihood function is defined through a Dirac delta function $\delta(\cdot)$:

$$h(\boldsymbol{x}|\boldsymbol{y}, m, \theta_m) = \prod_{j=1}^{k} \delta(y_j - x_j), \tag{31}$$

thus with $\Psi_m = \Theta_m$ the likelihood function in Expression (2) becomes

$$h(\boldsymbol{x}|m, \theta_m) = \prod_{j=1}^{k} f(x_j|m, \theta_m). \tag{32}$$

Under the two suggested models in Expressions (29) and (30), this likelihood function can be expressed as follows:

$$\begin{aligned} h(\boldsymbol{x}|m_1, \theta_1) &= (\prod_{j=1}^{k} x_j)^{-1} \cdot \theta_1^k \exp\{-\theta_1 \sum_{j=1}^{k} (\ln x_j - \ln x_0)\}, \\ h(\boldsymbol{x}|m_2, \theta_2) &= \theta_2^k \exp\{-\theta_2 \sum_{j=1}^{k} (x_j - x_0)\}, \end{aligned}$$

where $x_0 = y_0$. Both densities belong to an exponential family, and are given on the form

$$h(\boldsymbol{x}|m, \theta_m) = c_m(\boldsymbol{x})\, \theta_m^k \exp\{-\theta_m t_m(\boldsymbol{x})\}, \tag{33}$$

where the functions $c_m(\boldsymbol{x})$ and $t_m(\boldsymbol{x})$ are given as $c_1(\boldsymbol{x}) = (\prod_{j=1}^{k} x_j)^{-1}$, $c_2(\boldsymbol{x}) = 1$, $t_1(\boldsymbol{x}) = \sum_{j=1}^{k} (\ln x_j - \ln x_0)$ and $t_2(\boldsymbol{x}) = \sum_{j=1}^{k} (x_j - x_0)$. From Expression (6) the marginal pdf of $\boldsymbol{X}$ under model m is given as

$$f(\boldsymbol{x}|m) = c_m(\boldsymbol{x}) \int \theta_m^k \exp\{-\theta_m t_m(\boldsymbol{x})\} f(\theta_m|m)\, d\theta_m. \tag{34}$$

This pdf can be found analytically if conjugate priors of the exponential family are chosen, for example gamma distributions with hyper-parameters α_m and β_m, with pdf

$$f(\theta_m|m) = \Gamma(\alpha_m)^{-1} \beta_m^{-\alpha_m} \theta_m^{\alpha_m - 1} \exp(-\theta_m/\beta_m), \qquad \theta_m \geq 0,\ \alpha_m, \beta_m > 0 \tag{35}$$

and mean and variance $\mathrm{E}[\Theta_m|m] = \alpha_m \beta_m$, $\mathrm{Var}[\Theta_m|m] = \alpha_m \beta_m^2$. The marginal pdf in (34) is then found to be

$$\begin{aligned} f(\boldsymbol{x}|m) &= c_m(\boldsymbol{x}) \Gamma(\alpha_m)^{-1} \beta_m^{-\alpha_m} \int_0^\infty \theta_m^{\alpha_m + k - 1} \exp\{-\theta_m(\beta_m^{-1} + t_m(\boldsymbol{x}))\}\, d\theta_m \\ &= c_m(\boldsymbol{x}) \Gamma(\alpha_m)^{-1} \Gamma(\alpha_m + k) \beta_m^{-\alpha_m} (\beta_m^{-1} + t_m(\boldsymbol{x}))^{-(\alpha_m + k)}, \end{aligned} \tag{36}$$

and Bayes factor is given as

$$B_{12}(\boldsymbol{x}) = (\prod_{j=1}^{k} x_j)^{-1} \frac{\beta_2^{\alpha_2}(\beta_2^{-1} + \sum_{j=1}^{k}(x_j - x_0))^{\alpha_2+k}}{\beta_1^{\alpha_1}(\beta_1^{-1} + \sum_{j=1}^{k}(\ln x_j - \ln x_0))^{\alpha_1+k}} \cdot \frac{\Gamma(\alpha_2)\Gamma(\alpha_1+k)}{\Gamma(\alpha_1)\Gamma(\alpha_2+k)}. \tag{37}$$

The posterior distribution of Θ_m under model m is a gamma distribution with parameters $(\alpha_m + k)$ and $(\beta_m^{-1} + t_m(\boldsymbol{x}))^{-1}$, and mean and variance

$$\mathrm{E}\left[\Theta_m|\boldsymbol{x}, m\right] = \frac{\beta_m(\alpha_m+k)}{1+\beta_m t_m(\boldsymbol{x})}, \qquad \mathrm{Var}\left[\Theta_m|\boldsymbol{x}, m\right] = \frac{\beta_m^2(\alpha_m+k)}{(1+\beta_m t_m(\boldsymbol{x}))^2}. \tag{38}$$

Similar calculations can be executed using an ordinary exponential distribution with no lower bound, replacing the function $t_2(\boldsymbol{x})$ with $t_2(\boldsymbol{x}) = \sum_{j=1}^{k} x_j$.

4.2 Non-exact observations

Observations of fault sizes are in general often contaminated with measurement errors. Furthermore, only a limited number $N \leq K$ of the faults are observed. For each variable Y_j there is a probability $p(Y_j)$ of observing the fault. The vector $\boldsymbol{S} = (S_1, S_2, \ldots, S_K)$ of indicator variables describes which variables that are observed, where $S_j = 1$ if variable Y_j is observed and $S_j = 0$ otherwise. The number of observed variables is thus $N = \sum_{j=1}^{K} S_j$. If a variable Y_j is observed, the likelihood of the measured value, denoted Z_j, is described by the function $h(z_j|y_j)$. The complete set of observations is denoted $\boldsymbol{X} = (\boldsymbol{S}, \boldsymbol{Z})$.

The unknown variable K has the same interpretation under both models, being the total number of faults. However, the posterior pdfs of K can differ significantly between the two models, and instead of handling a problem of bimodality of K in the posterior, auxiliary variables K_1 and K_2 are introduced to represent the number of faults under the two models. The unknown dimension K_m of $\boldsymbol{Y}$ under model m is thus included in the model parameter $\Psi_m = (K_m, \Theta_m)$. Furthermore, depending on how prior knowledge about K_m has been obtained, the prior pdfs of K_m may be chosen differently under the two models. This is discussed by Borgos (2000). The sample spaces of K_1 and K_2 are given as $\Omega_{k_1} = \Omega_{k_2} = \{0, 1, 2, \ldots\}$. However, since both K_1 and K_2 must satisfy $K_m \geq N$, given the observations, the likelihood function $h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m)$ has value zero for $K_m < N$. For $K_m \geq N$, the likelihood function of the observations $\boldsymbol{X}$ is given as

$$h(\boldsymbol{x}|\boldsymbol{y}, m, \psi_m) = \prod_{j=1}^{k_m} p(y_j)^{s_j}(1-p(y_j))^{1-s_j} \prod_{j;s_j=1} h(z_j|y_j). \tag{39}$$

The reduced likelihood function in Expression (2) is found to be

$$\begin{aligned} h(\boldsymbol{x}|m,\psi_m) &= \prod_{j;s_j=1} \int h(z_j|y_j)p(y_j)f(y_j|m,\theta_m)\,dy_j \prod_{j;s_j=0} \int (1-p(y_j))f(y_j|m,\theta_m)\,dy_j \\ &= (1-p(m,\theta_m))^{k_m-n} \prod_{i=1}^{n} h(z_i|m,\theta_m), \end{aligned} \tag{40}$$

where

$$p(m,\theta_m) = \int p(y)f(y|m,\theta_m)\,dy, \qquad h(z|m,\theta_m) = \int h(z|y)p(y)f(y|m,\theta_m)\,dy. \tag{41}$$

For simplicity, the observed values $\boldsymbol{z}$ are renumbered from $\{j; s_j = 1\}$ to $i = 1,\ldots,n$, where $n = \sum_{j=1}^{k_m} s_j$. In the work by Borgos (2000) the probability $p(y)$ is an monotonely increasing, piecewise linear function and $h(z|y)$ is a uniform pdf on a closed interval determined by y, thus the integrals in Expression (41) are analytically tractable.

The model choice problem is solved using the McMC algorithm described below. When a model m has been chosen and the corresponding model parameters $\Psi_m = (K_m, \Theta_m)$ estimated, realizations of $\boldsymbol{Y}$ from $f(\boldsymbol{y}|\boldsymbol{x},m,\psi_m)$, Expression (4), can be obtained by rejection sampling. The posterior pdf of $\boldsymbol{Y}$ can be expressed as

$$f(\boldsymbol{y}|\boldsymbol{x},m,\psi_m) = \prod_{j;s_j=1} \frac{h(z_j|y_j)p(y_j)f(y_j|m,\theta_m)}{h(z_j|m,\theta_m)} \prod_{j;s_j=0} \frac{(1-p(y_j))f(y_j|m,\theta_m)}{1-p(m,\theta_m)},$$

see Expression (40). The pdf $f(y_j|m,\theta_m)$ can be used as a proposal distribution for all j, and the sample y_j is accepted with probability $h(z_j|y_j)p(y_j)/\max_y\{h(z_j|y)p(y)\}$ if $s_j = 1$ and with probability $(1-p(y_j))/\max_y\{1-p(y)\}$ if $s_j = 0$. For the piecewise linear function $p(y)$ and the uniform pdf $h(z|y)$ used by Borgos (2000), the maximum values are easily obtained.

4.3 McMC algorithm

Algorithm 1 is used to solve the model choice problem for fault size distributions, and to estimate the parameters of the preferred model. The algorithm is constructed for the general case of non-exact observations. The target distribution in Expression (8) is now given as

$$\begin{aligned} &\pi(m,\psi_m,\psi_{m_*}|\boldsymbol{x}) = \\ &\quad \text{const} \times (1-p(m,\theta_m))^{k_m-n} \prod_{i=1}^{n} h(z_i|m,\theta_m) f(k_m,\theta_m|m)\pi(k_{m_*},\theta_{m_*}|m)\pi(m), \end{aligned} \tag{42}$$

where $\boldsymbol{X} = (\boldsymbol{S}, \boldsymbol{Z})$ and $\Psi_m = (K_m, \Theta_m)$. The pseudoprior of Ψ_m, based on realizations $\{\psi_m^j\}$, $j = 1, \ldots, r$ from the posterior pdf $f(\psi_m|\boldsymbol{x}, m)$, is given by the kernel density estimate

$$\pi(\psi_m|m_*) = \frac{1}{r}\sum_{j=1}^{r} g_k(k_m|k_m^j, \lambda_m) g_\theta(\theta_m|\theta_m^j, \sigma_m). \tag{43}$$

A truncated Gaussian kernel with variance σ_m^2 is used for Θ_m, conditioned on $\Theta_m \in \Omega_{\theta_m}$. A Poisson distribution with parameter λ_m is used for the absolute value of $K_m - k_m^j$, with equal probability of positive and negative difference. If $k_m < k_m^j$, conditioning on $K_m \in \Omega_{k_m}$ is included in the kernel density. The parameters σ_m and λ_m are chosen as

$$\sigma_m = (4/3)^{1/5}\,\widehat{\mathrm{std}}(\Theta_m)\, r^{-1/5}, \qquad \lambda_m = (4/3)^{1/5}\,\widehat{\mathrm{std}}(K_m)\, r^{-1/5},$$

see Silverman (1986), where the standard deviations of Θ_m and K_m are estimated based on the realizations $\{\psi_m^j\}$, $j = 1, \ldots, r$.

There are three update steps within one iteration of Algorithm 1, updating M, Ψ_m and Ψ_{m_*} separately. When M is updated, the transition density is chosen as $q(m'|m) = \mathbf{1}[m' \neq m]$, thus $m' = m_*$ is always suggested. The acceptance probability is given in Expression (11), where $q(m_*|m) = q(m|m_*) = 1$ cancels. New values of Θ_m and K_m are suggested independently, using a transition kernel

$$q(\psi_m'|\psi_m, m) = q(\theta_m'|\theta_m, m) q(k_m'|k_m, m). \tag{44}$$

A new value θ_m' is suggested from a Gaussian distribution with mean θ_m and variance $\tau_m = 0.05\,\mathrm{E}\,[\Theta_m|m]$, truncated at $\theta_m \pm 5\tau_m$. Thus $q(\theta_m'|\theta_m, m) = q(\theta_m|\theta_m', m)$. If $\theta_m' \leq 0$, i.e., $\theta_m' \notin \Omega_{\theta_m}$, the acceptance probability in Expression (12) is replaced by $\alpha(\psi_{m_*}'|\psi_{m_*}, m) = 0$ and the suggested values $\psi_m' = (\theta_m', k_m')$ are not accepted, regardless of k_m and k_m'.

The current number k_m is suggested changed to either $k_m + 1$ or $k_m - 1$, choosing the first alternative with probability $\beta(k_m)$, the second with probability $1 - \beta(k_m)$. Since $K_m \geq n$, the function value $\beta(n)$ is chosen as $\beta(n) = 1$. An increase from k_m to $k_m + 1$ corresponds to adding a new element to the vector $\boldsymbol{Y}$ of faults, where the new element has $k_m + 1$ possible index numbers. A decrease from k_m to $k_m - 1$ corresponds to removing one of the unobserved faults, that is, removing a Y_j where $S_j = 0$. There are then $k_m - n$ possible indexes j to choose from. The two alternative transition kernels of K_m are thus given as

$$q(k_m + 1|k_m, m) = \frac{\beta(k_m)}{k_m + 1}, \qquad q(k_m - 1|k_m, m) = \frac{1 - \beta(k_m)}{k_m - n}.$$

When $\theta_m' > 0$, the acceptance probability for the suggested change in Ψ_m is given in Expression (12). The pseudoprior (43) can easily be sampled from, and is used as transition kernel for Ψ_{m_*}. Thus $q(\psi_{m_*}'|\psi_{m_*}, m) = \pi(\psi_{m_*}'|m)$ and the acceptance probability in Expression (13) is $\alpha(\psi_{m_*}'|\psi_{m_*}, m) = 1$.

The McMC algorithm can easily be simplified to deal with the case of exact observations. The likelihood function from Expression (40) is replaced by the function in Expression (32), and the number of faults, K, is no longer treated as a stochastic variable. This has implications on the pseudoprior and the transition kernels of Ψ_m and Ψ_{m_*}, where all terms involving K are removed.

5 Evaluation of the McMC Algorithm

The behavior of Bayes factor and the performance of Algorithm 1 are examined by considering the special case of a Pareto distribution and an exponential distribution as competing models, see Section 4. Bayes factor can be found analytically in the case of exact observations, which permits a comparison of the estimated Bayes factor with the analytical value in Expression (37). The posterior means of the model parameters can be estimated based on samples from the McMC algorithm, see Expression (18), and can be compared to the exact posterior means in Expression (38).

Consider the case of exact observations. In some of the simulation studies presented below a true model m is chosen, with a corresponding model parameter θ_m, and a set of observations $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$ is generated from $f(\boldsymbol{x}|m, \theta_m)$ using the distributions presented in Section 4. Bayes factor is then calculated for a selection of values of the hyper-parameters in the prior pdf of Θ_{m_*}, $m_* \neq m$. Gamma priors are used, Expression (35), with a common constant shape parameter $\alpha_1 = \alpha_2 = 4$ for both parameters Θ_1 and Θ_2. For the true model the scale parameter is chosen as $\beta_m = \theta_m/\alpha_m$, thus the chosen parameter value coincides with the prior mean $\mathrm{E}[\Theta_m|m] = \alpha_m\beta_m$. For the other model the scale parameter is given as $\beta_{m_*} = \mathrm{E}[\Theta_{m_*}|m_*]/\alpha_{m_*}$, where a number of values of the prior mean $\mathrm{E}[\Theta_{m_*}|m_*]$ is used to explore Bayes factor. Before the simulation study is carried out, the range of values of $\mathrm{E}[\Theta_{m_*}|m_*]$ to be used was determined by a preliminary simulation study.

To study the convergence of Algorithm 1, both the convergence of Ψ_m and and the mixing of M should be explored. The convergence of the algorithm can be examined by plotting different functions of the variables for realizations from the simulations. Several runs of the algorithm should be compared, using different initial states, to reveal any dependence on the starting point. The burn-in period of the algorithm is used to secure convergence of Ψ_m to the posterior distribution under model m, and M is of no interest. During these iterations independent simulations of Ψ_m and Ψ_{m_*} are performed with their respective posterior distributions as target distributions, using transition kernels $q(\psi'_m|\psi_m, m)$ and $q(\psi'_{m_*}|\psi_{m_*}, m_*)$. After the burn-in a continued sampling of Ψ_m and Ψ_{m_*} from their respective posterior distributions is performed, giving realizations from which the pseudopriors can be defined. Finally, the McMC algorithm is run with initial states $\psi_m(0)$ chosen as the final realizations of the preliminary runs and $m(0)$ uniformly drawn from Ω_m. The mixing properties of the Markov chain is then examined based on the realizations of the model

indicator M.

In Section 5.1 the behavior of Bayes factor is studied through randomization of $\boldsymbol{X}$, while in Section 5.2 the performance of Algorithm 1 is explored in the case of exact observations, using a single dataset $\boldsymbol{X}$ generated from $f(\boldsymbol{x}|m, \theta_m)$. In Section 5.3 an example with non-exact observations is presented.

5.1 Behavior of analytical Bayes factor for exact observations

The behavior of the analytical Bayes factor in the case of exact observations, see Expression (37), is studied by generating 25 000 realizations of the vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$ from the Pareto distribution with parameter $\Theta_1 = 1.0$ and lower limit $x_0 = 2$. For each set of observations, Bayes factor is calculated for a selection of prior mean values $\mathrm{E}[\Theta_2|m_2]$. Model 1 is known to be the correct model, thus Bayes factor should be $B_{12} > 1$. However, due to natural variations in the generated samples, the wrong model, model 2, in some cases fits the data better than the true model. Table 1 reports the results for sample sizes of $k = 10$ and $k = 100$, listing the mean values of Bayes factor based on the 25 000 samples. The geometric mean and the median are also reported. The empirical probability $P(B_{12} < 1)$ gives the proportion of samples where model 2 is favored. Within the range of prior mean values of Θ_2 reported in Table 1, Bayes factor can be $B_{12} < 1$ for as much as $25\% - 30\%$ of the realizations for sample sizes of $k = 10$. As the size of the dataset increases, the proportion of wrong conclusions decreases, and for $k = 100$ Bayes factor favors model 1 in allmost all simulations.

Table 2 lists the result of a similar simulation study, where 25 000 samples of size $k = 10$ and $k = 100$ are generated from the exponential distribution with parameter $\Theta_2 = 0.2$ and lower limit $x_0 = 2$. Bayes factor is evaluated for different values of the prior mean $\mathrm{E}[\Theta_1|m_1]$.

The results of Tables 1 and 2 show that for exact observations from either the Pareto or the exponential distribution, there is usually no doubt which model is the right one, even for fairly small datasets. In the case of non-exact observations however, there may be greater uncertainties about which model is most likely to be the correct one.

5.2 Performance of the algorithm for exact observations

The performance of Algorithm 1 in the case of exact observations is studied by generating a dataset $\boldsymbol{X}$ consisting of $k = 10$ observations from model 1, a Pareto distribution with parameter $\Theta_1 = 1.0$ and lower limit $x_0 = 2$. For this particular dataset, the posterior mean of Θ_1 is $\mathrm{E}[\Theta_1|\boldsymbol{x}, m_1] = 0.732$, see Expression (38), while the prior mean is $\mathrm{E}[\Theta_1|m_2] = 1.0$. Inference is based on realizations from the algorithm, repeating the procedure using

		$\mathrm{E}[\Theta_2\|m_2]$				
		0.05	0.10	0.20	0.30	0.40
B_{12}	arithmetic mean	$\mathcal{O}(10^{41})$	$\mathcal{O}(10^{42})$	$\mathcal{O}(10^{43})$	$\mathcal{O}(10^{44})$	$\mathcal{O}(10^{45})$
	geometric mean	564.36	100.52	69.80	97.61	151.76
	median	200.77	18.26	5.67	6.04	9.12
$\hat{P}(B_{12} < 1)$		0.037	0.129	0.248	0.282	0.268

		$\mathrm{E}[\Theta_2\|m_2]$				
		0.05	0.10	0.15	0.20	0.25
B_{12}	arithmetic mean	$\mathcal{O}(10^{300})$	$\mathcal{O}(10^{301})$	$\mathcal{O}(10^{302})$	$\mathcal{O}(10^{302})$	$\mathcal{O}(10^{303})$
	geometric mean	$\mathcal{O}(10^{37})$	$\mathcal{O}(10^{37})$	$\mathcal{O}(10^{37})$	$\mathcal{O}(10^{37})$	$\mathcal{O}(10^{37})$
	median	$\mathcal{O}(10^{29})$	$\mathcal{O}(10^{28})$	$\mathcal{O}(10^{28})$	$\mathcal{O}(10^{29})$	$\mathcal{O}(10^{29})$
$\hat{P}(B_{12} < 1)$		0	0.00012	0.00020	0.00024	0.00024

Table 1: Variation in Bayes factor for 25 000 realizations generated from the Pareto distribution with parameter $\Theta_1 = 1.0$, using different prior means of the parameter Θ_2 in the competing exponential distribution. In the upper table $k = 10$, in the lower table $k = 100$. Arithmetic mean, geometric mean and median of the exact values of Bayes factor are reported, together with the proportions of Bayes factors of values below 1. $\mathcal{O}(x)$ gives the order of magnitude.

		$\mathrm{E}[\Theta_1\|m_1]$				
		0.5	0.75	1.00	1.25	1.50
B_{12}	arithmetic mean	0.176	0.376	0.488	0.516	0.494
	geometric mean	0.067	0.120	0.133	0.123	0.105
	median	0.061	0.112	0.125	0.116	0.098
$\hat{P}(B_{12} < 1)$		0.978	0.939	0.917	0.915	0.920

		$\mathrm{E}[\Theta_1\|m_1]$				
		0.85	0.95	1.00	1.05	1.10
B_{12}	arithmetic mean	0.00538	0.00622	0.00651	0.00672	0.00687
	geometric mean	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$
	median	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-8})$
$\hat{P}(B_{12} < 1)$		0.9998	0.9998	0.9998	0.9998	0.9998

Table 2: Variation in Bayes factor for 25 000 realizations generated from the exponential distribution with parameter $\Theta_2 = 0.2$, using different prior means of the parameter Θ_1 in the competing Pareto distribution. In the upper table $k = 10$, in the lower table $k = 100$. Arithmetic mean, geometric mean and median of the exact values of Bayes factor are reported, together with the proportions of Bayes factors of values below 1. $\mathcal{O}(x)$ gives the order of magnitude.

different values of $\mathrm{E}[\Theta_2|m_2]$. The algorithm is first run to generate realizations of Θ_m, $m = 1, 2$, from their respective posterior distributions $f(\theta_m|\boldsymbol{x}, m)$. These realizations are used to define pseudopriors, and each realization is separated by 10 iterations, after a burn-in of 5000 iterations. The last realizations of Θ_1 and Θ_2 are used as initial states $\theta_1(0)$ and $\theta_2(0)$ in the final run of Algorithm 1, while an initial state $m(0)$ is chosen uniformly on Ω_m. The number of iterations is 50 000. All realizations of M are used to estimate Bayes factor and to find a confidence interval, while inference about the model parameters are based on every $10th$ realization of (Θ_1, Θ_2).

The convergence of Θ_1 and Θ_2 during a burn-in period is illustrated in Figure 1, using $\mathrm{E}[\Theta_2|m_2] = 0.1$. The figure shows realizations of Θ_m and the logarithmic values of the pdfs $f(\boldsymbol{x}|m, \theta_m)$ from 5000 iterations. Four different starting points for each parameter are used, chosen as $10^{-6}\,\mathrm{E}[\Theta_m|m]$, $\mathrm{E}[\Theta_m|m]$, $10\,\mathrm{E}[\Theta_m|m]$ and $25\,\mathrm{E}[\Theta_m|m]$. The posterior mean is close to the prior mean for both model parameters in the case studied in Figure 1, thus with the prior means as initial states convergence is soon established. Starting close to zero, the McMC algorithm also converges rapidly. With a starting point an order of magnitude larger than the posterior mean, a larger number of iterations is required before the chain converges, but in both examples convergence is clearly reached within 5000 iterations. The plots of $\log f(\boldsymbol{x}|m, \theta_m)$ support the interpretation of the convergence. The speed of convergence in the case of exact observations depend on the size of the update steps, defined through the transition kernel $q(\theta'_m|\theta_m, m)$, see Section 4.3.

Figure 2 shows an output from the algorithm, based on 50 000 iterations, with $\mathrm{E}[\Theta_2|m_2] = 0.1$. 5000 burn-in iterations are used, and the pseudopriors are based on 5000 realizations of Θ_1 and Θ_2, separated by 10 iterations. The numerical values of the estimates from this run of the algorithm are reported below, in Table 3. The plots of $\pi(m, \theta_1, \theta_2|\boldsymbol{x})$, θ_1 and θ_2 all seem stable without any trends, see Figure 2a–c. Based on the plot of the convergence of the estimate of Bayes factor, Figure 2d, it seems that in this case 50 000 iterations is sufficient to give a reliable estimate. The samples of M are not plotted, but from the acceptance rate and autocorrelation given in Table 3 it is clear that the mixing of M is satisfactory. The histograms of Θ_m in Figure 2e,f show that in this example the posterior pdfs are narrower than the prior pdfs, but the posterior pdfs are not considerably skewed compared to the priors. Figures 2g,h show the function

$$S(x|m, \theta_m) = \mathrm{P}\{X \geq x|m, \theta_m\} = \int_x^\infty f(x|m, \theta_m)\, dx, \tag{45}$$

using the posterior mean estimates as model parameters, together with the empirical estimate

$$\hat{S}(x|m, \theta_m) = \frac{1}{k} \sum_{j=1}^{k} \mathbf{1}[x_j \geq x].$$

Under model 1 the transformation $(\log x, \log S(x|m, \theta_m))$ gives a linear plot, while under model 2 linearity is obtained for $(x, \log S(x|m, \theta_m))$. The comparisons of the parametric

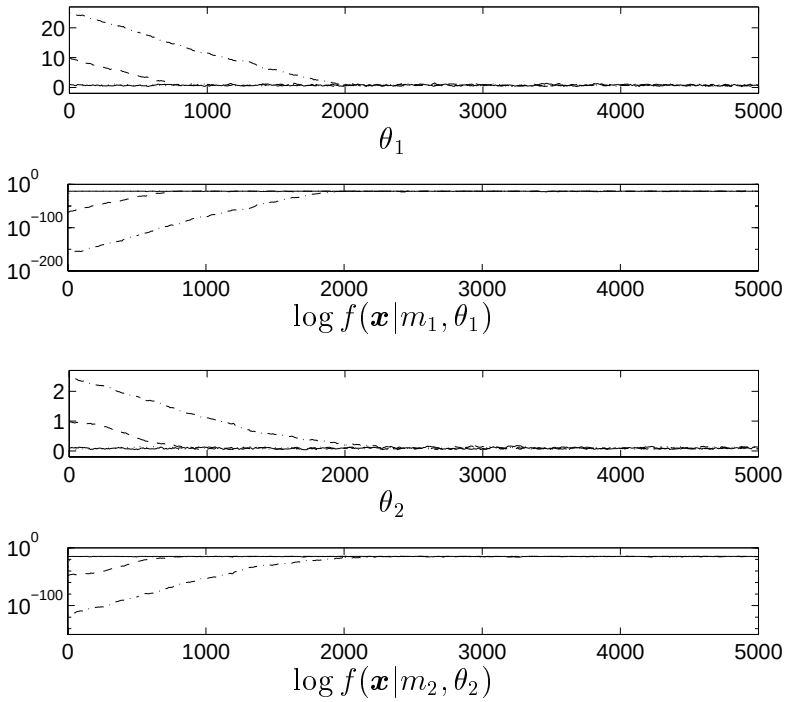


Figure 1: Realizations of Θ_1 and Θ_2 and the corresponding $\log f(\boldsymbol{x}|m, \theta_m)$ from 5000 burn-in iterations of Algorithm 1. Solid lines correspond to runs with initial values $\theta_m(0) = \mathrm{E}[\Theta_m|m]$. Dotted lines have initial values close to zero and are hardly distinguishable from the solid lines. For the dashed lines and dashed/dotted lines the initial values are $10\,\mathrm{E}[\Theta_m|m]$ and $25\,\mathrm{E}[\Theta_m|m]$ respectively.

estimates of $S(x|m, \theta_m)$ and the empirical estimates give a visual illustration of how well the observations are fitted by each of the two models.

Repeated simulations are run with a varying value of $\mathrm{E}[\Theta_2|m_2]$, and the results are listed in Table 3. The preliminary estimate $\tilde{B}_{12}(\boldsymbol{x})$ in Expression (26) based on $L = 1$ parameter value, the final estimate $\hat{B}_{12}(\boldsymbol{x})$ from Expression (16), and an approximate 95% confidence interval, Expression (27), are compared with the analytical value of Bayes factor. The acceptance rate of proposed changes of model and the estimated autocorrelation $\rho(1)$ at lag 1 are also reported. Estimated posterior mean values $\hat{\mathrm{E}}[\Theta_m|\boldsymbol{x}, m]$ of the model parameters, Expression (18), are compared to the true posterior means. The estimated standard deviation of the estimator $\hat{\mathrm{E}}[\Theta_m|\boldsymbol{x}, m]$ is given in brackets, obtained as $(\hat{\mathrm{Var}}[\Theta_m|\boldsymbol{x}, m]/n_m)^{1/2}$, where n_m is the number of realizations from model m, $\{\theta_m^j|m^j = m\}$.

For the dataset used in this simulations study, Bayes factor obtains a minimum as a function of $\beta_2 = \mathrm{E}[\Theta_2|m_2]/\alpha_2$ when $\mathrm{E}[\Theta_2|m_2] \approx 0.10$. When $\mathrm{E}[\Theta_2|m_2]$ is close to this value Bayes factor is $B_{12} < 1$, which means that model 2 actually fits the data better. From the results in Table 1, it is clear that this is not a rare event for small sample sizes. As $\mathrm{E}[\Theta_2|m_2]$

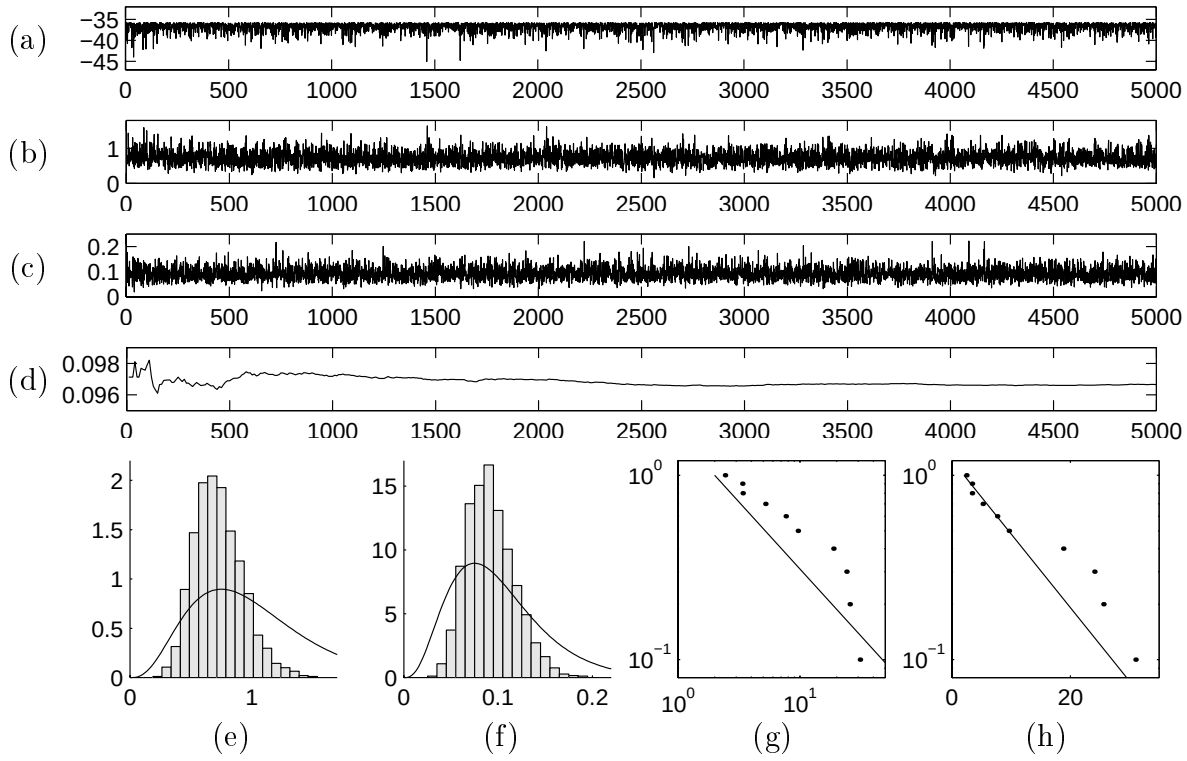


Figure 2: Output from a run of Algorithm 1 with exact observations. 5000 samples of (M, θ_1, θ_2) are generated from $\pi(m, \theta_1, \theta_2|\boldsymbol{x})$, separated by 10 iterations. The first plots are trace plots of (a) $\log \pi(m, \theta_1, \theta_2|\boldsymbol{x})$, (b) θ_1, (c) θ_2 and (d) $\hat{B}_{12}(\boldsymbol{x})$. Histograms of (e) $\{\theta_1|m = m_1\}$ and (f) $\{\theta_2|m = m_2\}$ are plotted, comparing with the corresponding prior pdfs (solid lines). The last two figures shows $S(x|m, \hat{\theta}_m)$ (solid lines), see Expression (45), together with empirical estimates. The functions are plotted on log-log scale under model 1, figure (g), and with logarithmic vertical axis under model 2, figure (h).

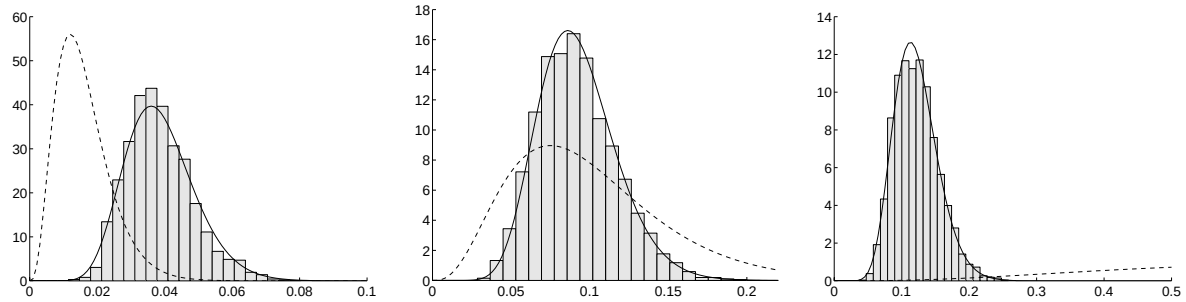


Figure 3: Histograms of 5000 samples from the posterior distribution of $[\Theta_2|m_2]$, compared with the posterior pdf (solid line) and prior pdf (dashed lines).

deviates from 0.1, the analytical value of Bayes factor increases and eventually $B_{12} > 1$ and the correct model is favored. The estimate of Bayes factor captures this increase both for increasing and decreasing values of $\mathrm{E}[\Theta_2|m_2]$, and the analytical Bayes factor is covered by the estimated confidence interval in all simulations reported in Table 3. The estimated autocorrelation shows a clear negative correlation between succeeding realizations of M, thus the estimated confidence intervals are likely to be too conservative. The estimated posterior mean values of the parameters lie within two estimated standard deviations from the exact posterior mean in most cases, and Figure 3 shows how the histograms of realizations of $[\Theta_2|m_2]$ coincide with the posterior pdf. A comparison with the prior pdf is also included in the figure.

A similar simulation study is run by generating a sample of size $k = 10$ from model 2, an exponential distribution with parameter $\Theta_2 = 0.1$ and lower limit $x_0 = 2$. Bayes factor for this dataset is found, varying the prior mean $\mathrm{E}[\Theta_1|m_1]$, and the results are listed in Table 4. Once more the algorithm seems to give satisfactory results of both estimated Bayes factor and model parameters.

The results reported in Tables 3 and 4 illustrate the performance of the algorithm in the simplest case of exact observations and a single model parameter. Only two datasets are presented in the simulation study above, one from each distribution (29) and (30), but repeated simulations are expected to give similar results. Small datasets are chosen to illustrate the performance of the algorithm, since Bayes factor is then likely to have a value not too far from 1, see Tables 1 and 2. In this situation it is important to be able to estimate Bayes factor accurately, to give the right conclusion of the model choice problem.

$\mathrm{E}[\Theta_2\|m_2]$	$B_{12}(\boldsymbol{x})$	$\tilde{B}_{12}(\boldsymbol{x})$	$\hat{B}_{12}(\boldsymbol{x})$	95% CI	rate	$\hat{\rho}(1)$
0.010	249.222	290.983	248.174	(243.848,252.577)	0.895	-0.80
0.016	12.609	13.930	12.652	(12.432,12.876)	0.929	-0.86
0.025	1.358	1.311	1.356	(1.332,1.380)	0.945	-0.89
0.040	0.268	0.263	0.268	(0.263,0.272)	0.954	-0.91
0.063	0.116	0.112	0.115	(0.113,0.118)	0.971	-0.94
0.10	0.097	0.093	0.097	(0.095,0.098)	0.964	-0.93
0.16	0.147	0.151	0.148	(0.145,0.150)	0.970	-0.94
0.25	0.336	0.339	0.335	(0.329,0.341)	0.970	-0.94
0.40	1.120	1.122	1.119	(1.100,1.139)	0.976	-0.95
0.63	4.489	4.197	4.499	(4.421,4.579)	0.961	-0.92
1.0	21.475	20.396	21.465	(21.092,21.844)	0.967	-0.93
1.6	117.124	125.897	117.438	(115.396,119.517)	0.961	-0.92

$\mathrm{E}[\Theta_2\|m_2]$	$\mathrm{E}[\Theta_2\|\boldsymbol{x},m_2]$	$\hat{\mathrm{E}}[\Theta_2\|\boldsymbol{x},m_2]$	$\hat{\mathrm{E}}[\Theta_1\|\boldsymbol{x},m_1]$
0.010	0.0274	0.0274 (0.0001)	0.723 (0.004)
0.016	0.0388	0.0386 (0.0002)	0.731 (0.004)
0.025	0.0516	0.0515 (0.0003)	0.744 (0.004)
0.040	0.0663	0.0666 (0.0004)	0.731 (0.004)
0.063	0.0802	0.0803 (0.0004)	0.733 (0.004)
0.10	0.0927	0.0919 (0.0005)	0.727 (0.004)
0.16	0.1029	0.1025 (0.0005)	0.739 (0.004)
0.25	0.1102	0.1102 (0.0006)	0.731 (0.004)
0.40	0.1156	0.1149 (0.0006)	0.730 (0.004)
0.63	0.1192	0.1192 (0.0007)	0.733 (0.004)
1.0	0.1216	0.1222 (0.0007)	0.728 (0.004)
1.6	0.1233	0.1236 (0.0006)	0.724 (0.004)

Table 3: McMC results for a dataset of size $k = 10$ generated from the Pareto distribution with parameter $\Theta_1 = 1.0$, varying the prior mean of Θ_2. The exact value of $B_{12}(\boldsymbol{x})$ from Expression (37) is compared to the preliminary estimate $\tilde{B}_{12}(\boldsymbol{x})$, Expression (26), the final estimate $\hat{B}_{12}(\boldsymbol{x})$, Expression (16), and the confidence interval in Expression (27). The acceptance rate of suggested changes of model and the estimated autocorrelation $\rho(1)$ at lag 1 is reported. Estimates of model parameters are given with standard deviation of the estimates in brackets. In all simulations the prior and posterior mean of Θ_1 are $\mathrm{E}[\Theta_1|m_1] = 1.0$ and $\mathrm{E}[\Theta_1|\boldsymbol{x},m_1] = 0.732$.

$\mathrm{E}[\Theta_1\|m_1]$	$B_{12}(\boldsymbol{x})$	$\tilde{B}_{12}(\boldsymbol{x})$	$\hat{B}_{12}(\boldsymbol{x})$	95% CI	rate	$\hat{\rho}(1)$
0.10	0.000147	0.000139	0.000148	(0.000145,0.000151)	0.883	-0.77
0.16	0.00347	0.00361	0.00346	(0.00340,0.00352)	0.899	-0.80
0.25	0.0395	0.0430	0.0394	(0.0387,0.0401)	0.917	-0.84
0.40	0.253	0.309	0.252	(0.248,0.257)	0.888	-0.80
0.63	0.734	0.758	0.734	(0.721,0.747)	0.940	-0.88
1.00	1.075	1.032	1.074	(1.055,1.093)	0.948	-0.90
1.60	0.841	0.863	0.842	(0.827,0.857)	0.909	-0.82
2.50	0.416	0.398	0.417	(0.409,0.424)	0.952	-0.91
4.00	0.137	0.143	0.137	(0.135,0.139)	0.966	-0.93
6.30	0.0364	0.0355	0.0364	(0.0358,0.0371)	0.951	-0.90
10.00	0.00793	0.00763	0.00792	(0.00778,0.00806)	0.950	-0.90
16.00	0.00149	0.00141	0.00150	(0.00147,0.00152)	0.949	-0.90

$\mathrm{E}[\Theta_1\|m_1]$	$\mathrm{E}[\Theta_1\|\boldsymbol{x},m_1]$	$\hat{\mathrm{E}}[\Theta_1\|\boldsymbol{x},m_1]$	$\hat{\mathrm{E}}[\Theta_2\|\boldsymbol{x},m_2]$
0.10	0.282	0.280 (0.002)	0.1618 (0.0009)
0.16	0.405	0.403 (0.002)	0.1631 (0.0008)
0.25	0.547	0.550 (0.003)	0.1628 (0.0009)
0.40	0.714	0.713 (0.004)	0.1620 (0.0008)
0.63	0.877	0.874 (0.005)	0.1613 (0.0009)
1.00	1.029	1.027 (0.006)	0.1625 (0.0009)
1.60	1.156	1.151 (0.006)	0.1622 (0.0009)
2.50	1.249	1.252 (0.006)	0.1626 (0.0009)
4.00	1.319	1.321 (0.007)	0.1627 (0.0008)
6.30	1.366	1.373 (0.007)	0.1627 (0.0009)
10.00	1.398	1.398 (0.008)	0.1611 (0.0008)
16.00	1.420	1.426 (0.008)	0.1628 (0.0009)

Table 4: McMC results for a dataset of size $k = 10$ generated from the exponential distribution with parameter $\Theta_2 = 0.1$, varying the prior mean of Θ_1. The exact value of $B_{12}(\boldsymbol{x})$ from Expression (37) is compared to the preliminary estimate $\tilde{B}_{12}(\boldsymbol{x})$, Expression (26), the final estimate $\hat{B}_{12}(\boldsymbol{x})$, Expression (16), and the confidence interval in Expression (27). The acceptance rate of suggested changes of model and the estimated autocorrelation $\rho(1)$ at lag 1 is reported. Estimates of model parameters are given with standard deviation of the estimates in brackets. In all simulations the prior and posterior mean of Θ_2 are $\mathrm{E}[\Theta_2|m_2] = 0.1$ and $\mathrm{E}[\Theta_2|\boldsymbol{x},m_2] = 0.1626$.

5.3 Performance of the algorithm for non-exact observations

When observations of fault sizes contain sampling errors, and the dimension K of the variable $\boldsymbol{Y}$ is unknown, K is included in the McMC sampling. Borgos (2000) analyze a fault size dataset from the Gullfaks field, see Fossen and Rørnes (1996), and Algorithm 1 is used to estimate Bayes factor to decide if the data is best fitted by a Pareto or an exponential distribution. The observed number of faults is $n = 169$, while the total number is unknown. The convergence during burn-in iterations is first studied, and Figure 4 shows a convergence plot for the total number of faults under the two competing models, using four different initial states $k_m(0)$. Under both models, convergence is reached within 10 000 iterations. In this example the dimensions K_m under the two models are assigned equal prior distributions, Poisson distributions with mean 720. It is observed from Figure 4 that the two Markov chains stabilizes at different values, thus the posterior pdfs of K_m differ. Convergence plots for the model parameters are not included for these simulations, but also these parameters are found to reach convergence within 10 000 iterations.

Figure 5 shows an output from the McMC algorithm, based on 50 000 iterations. The numerical results of this and similar runs for the same dataset are presented by Borgos (2000). A burn-in of 10 000 iterations is used, and pseudopriors are based on 5000 realizations of Ψ_1 and Ψ_2, separated by 10 iterations. In this case the preliminary and final estimate of Bayes factor, Expressions (26) and (16), are $\tilde{B}_{12}(\boldsymbol{x}) = 0.01423$ and $\hat{B}_{12}(\boldsymbol{x}) = 0.00538$, the acceptance rate for suggested changes of M is 0.0136 and the estimated autocorrelation at lag 1 is $\hat{\rho}(1) = 0.966$. The rough estimate is based on $L = 100$ values ψ_m^j.

Based on the difference between the final estimate of Bayes factor and the rough estimate, it seems that the rough estimate is not as good as the case was for exact observations with only one unknown model parameter for each model. This also means that the pseudopriors $\pi(\psi_m|m_*)$ deviates to some extent from the corresponding posterior pdfs. This is not surprising, since 5000 realizations of Ψ_m is now used to estimate a two-dimensional

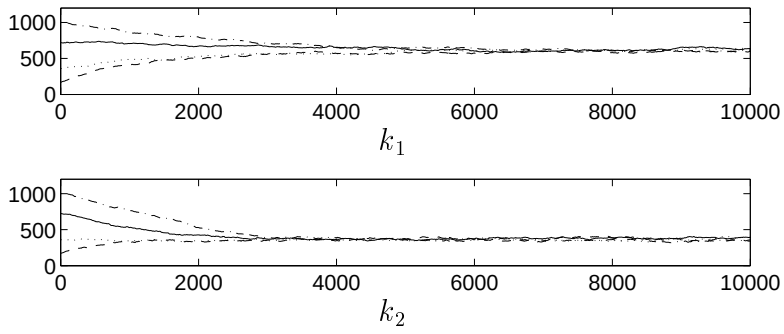


Figure 4: Realizations of K_1 and K_2 from 10 000 burn-in iterations of Algorithm 1, starting with different initial states $k_1(0)$ and $k_2(0)$.

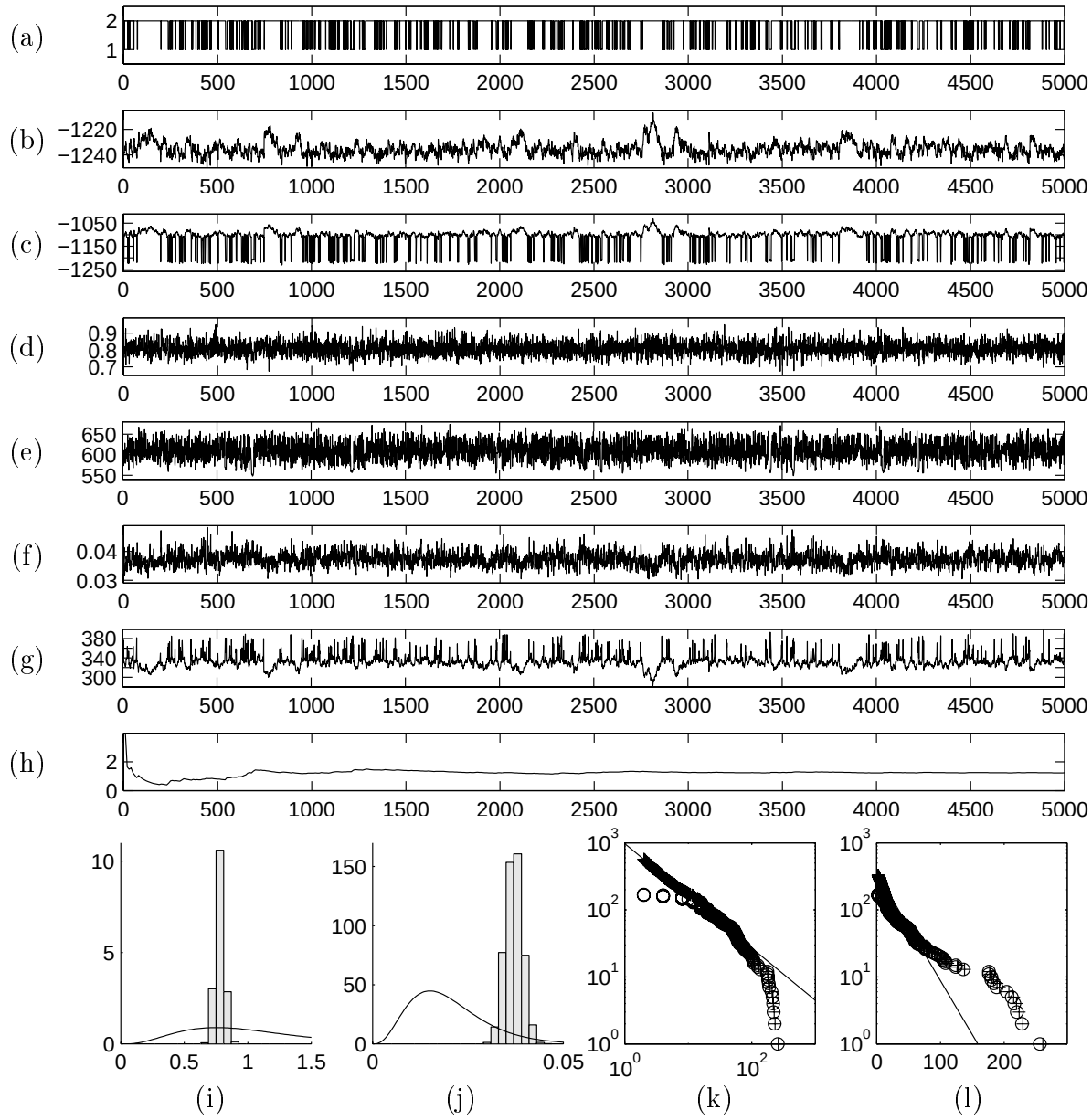


Figure 5: Output from a run of Algorithm 1 with non-exact observations. 5000 samples of (M, ψ_1, ψ_2) are generated from $\pi(m, \psi_1, \psi_2|\boldsymbol{x})$, separated by 10 iterations. The first plots are trace plots of (a) m, (b) $\log \pi(m, \psi_1, \psi_2|\boldsymbol{x})$, (c) $\log h(\boldsymbol{x}|m, \psi_m)$, (d) θ_1, (e) k_1, (f) θ_2 and (g) k_2 and (h) $\hat{B}_{12}(\boldsymbol{x})$. Histograms of (i) $\{\theta_1|m = m_1\}$ and (j) $\{\theta_2|m = m_2\}$ are plotted, comparing with the corresponding prior pdfs (solid lines). The last two figures shows $\hat{k}_m S(y|m, \hat{\theta}_m)$ (solid lines), see Expression (45), where $\hat{k}_m$ is the mean value of the dimension K_m obtained in the simulations. The function is compared to empirical estimates based on observations $\boldsymbol{x}$ and a realization $\boldsymbol{y}$ from $f(\boldsymbol{y}|\boldsymbol{x}, m, \hat{\psi}_m)$. The functions are plotted on log-log scale under model 1, figure (k), and with logarithmic vertical axis under model 2, figure (l).

joint distribution, while in the case of exact observations a univariate pdf was estimated based on the same number of samples. The mismatch between $\tilde{B}_{12}(\boldsymbol{x})$ and $B_{12}(\boldsymbol{x})$ gives an acceptance probability $\alpha(m'|m)$ far from 1, and instead of a negative autocorrelation a strong positive autocorrelation is obtained. The quality of the estimate $\tilde{B}_{12}(\boldsymbol{x})$ and the pseudopriors $\pi(\psi_m|m_*)$ could possibly be improved by generating a larger number of samples of Ψ_1 and Ψ_2 in the initial run of the McMC algorithm. However, this would require longer time for each iteration in the final run of the algorithm, since the time consumption for evaluating the pseudoprior in Expression (25) increases with the number of samples ψ_m^j from the initial run.

Although no negative autocorrelation in M is obtained, Figure 5a shows that the Markov chain still alternates satisfactorily between m_1 and m_2. The value of the target pdf in Expression (42), Figure 5b, the likelihood function (40), Figure 5c, and the samples of the parameters Ψ_1 and Ψ_2, Figure 5d–g, all indicate that convergence of the Markov chain is reached. From Figure 5h it seems that 50 000 iterations is sufficient to give a reliable estimate of Bayes factor. Figure 5k,l shows realizations of $\boldsymbol{Y}$ from the posterior pdfs $f(\boldsymbol{y}|\boldsymbol{x}, m, \hat{\psi}_m)$, comparing the estimated function $S(y|m, \hat{\theta}_m)$, Expression (45), with the empirical estimates obtained based on observations $\boldsymbol{x}$ and samples $\boldsymbol{y}$. The empirical functions are multiplied by the number of faults, n and $\hat{k}_m$ for $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, to illustrate at what scales the unobserved variables Y_j are sampled.

6 Conclusion

A model choice problem is considered, where inference is made based on observations affected by sampling errors. The underlying unknown variables can be divided into a high-dimensional part with an analytically tractable posterior distribution and a low-dimensional model parameter with an analytically non-tractable posterior. An McMC algorithm is constructed to estimate Bayes factor and make inference about the low-dimensional model parameters simultaneously. The unknown high-dimensional variable can afterwards be sampled from its posterior distribution. In order to obtain a completely specified target distribution of the McMC algorithm, the concept of pseudopriors introduced by Carlin and Chib (1995) is adopted. The freedom of choice for these pseudopriors is explored, under the major goal of reducing uncertainty in the estimated Bayes factor.

The performance of the algorithm is studied for a situation where Bayes factor is analytically tractable, and all estimates can be compared to the corresponding true, known values. The algorithm is observed to give good results in this case, where the observations are assumed to be exact and the model parameter is one dimensional. In this case a strong negative correlation between succeeding realizations of the model indicator can be produced, reducing the uncertainty in the estimated Bayes factor.

An example of non-exact data is provided. It is observed that as the dimension of the unknown parameters included in the McMC algorithm increases, it becomes more difficult to find good pseudopriors based on a relatively small number of samples of the parameters. This results in a slower mixing of M, and no negative autocorrelation is obtained. However, the model indicator M still changes relatively often throughout the simulations, and the algorithm seems to perform satisfactory also in this case.

If the posterior distribution of the unobserved high-dimensional parameter, or parts of it, were not analytically tractable, this variable could also have been included in the sampling algorithm. However, this expansion of the sample space is assumed to reduce the efficiency of the algorithm considerably.

References

Ackermann, R. V., Withjack, M. O. and Schlische, R. W. (1999). The geometric and statistical evolution of normal fault systems: An experimental study of the effects of mechanical layer thickness on scaling laws, submitted to *Journal of Structural Geology*.

Aitkin, M. (1991). Posterior Bayes Factors, *J. R. Statist. Soc. B* **53**(1): 111–142.

Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Association* **91**(433): 109–122.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion), *Statistical Science* **10**(1): 3–66.

Borgos, H. G. (2000). Model choice for fault size distribution, Part I of this thesis.

Carlin, B. P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods, *J. R. Statist. Soc. B* **57**(3): 473–484.

Carlin, P. and Polson, G. (1991). Inference for nonconjugate Bayesian Models using the Gibbs sampler, *The Canadian Journal of Statistics* **19**(4): 399–405.

Chib, S. (1995). Marginal Likelihood From the Gibbs Output, *Journal of the American Statistical Association* **90**(432): 1313–1321.

Cowie, P. A., Malinverno, A., Ryan, W. B. F. and Edwards, M. H. (1994). Quantitative fault studies on the East Pacific Rise: A comparison of sonar image techniques, *Journal of Geophysical Research* **99**(B8): 15205–15218.

Dellaportas, P., Forster, J. J. and Ntzoufras, I. (1998). On Bayesian Model and Variable Selection Using MCMC, *Technical Report No 40*, Department of Statistics, Athens University of Economics and Business. Submitted to *Statistics and Computing*.

DiCiccio, T. J., Kass, R. E., Raftery, A. and Wasserman, L. (1997). Computing Bayes Factors By Combining Simulation and Asymptotic Approximations, *Journal of the American Statistical Association* **92**(439): 903–915.

Evans, M. and Swartz, T. (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems, *Statistical Science* **10**(3): 254–272.

Fossen, H. and Rørnes, A. (1996). Properties of fault populations in the Gullfaks Field, northern North Sea, *Journal of Structural Geology* **18**(2/3): 179–190.

Gelman, A. and Meng, X.-L. (1998). Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling, *Statistical Science* **13**(2): 163–185.

George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* **88**(423): 881–889.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrica* **82**(4): 711–732.

Han, C. and Carlin, B. P. (2000). MCMC Methods for Computing Bayes Factors: A Comparative Review, *Research Report 2000-001*, Division of Biostatistics, University of Minnesota.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrica* **57**: 97–109.

JSG (1996). *Journal of Structural Geology*, Vol. 18, No. 2/3, Special Issue, Scaling Laws for Fault and Fracture Populations – Analyses and Applications.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association* **90**(430): 773–795.

O'Hagan, A. (1991). Contribution to the discussion of "Posterior Bayes Factors" (by M. Aitkin), *J. R. Statist. Soc. B* **53**(1): 136.

O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison, *J. R. Statist. Soc. B* **57**(1): 99–138.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Spryopoulos, C., Griffith, W. J., Scholz, C. H. and Shaw, B. E. (1999). Experimental evidence for different strain regimes of crack populations in a clay model, *Geophysical Research Letters* **26**(8): 1081–1084.

Part III

Practicalities of extrapolating 1D fault and fracture size-frequency distributions to higher dimensional samples

Practicalities of extrapolating 1D fault and fracture size-frequency distributions to higher dimensional samples

Hilde G. Borgos
Department of Mathematical Sciences
The Norwegian University of Science and Technology

Patience A. Cowie, Nancye H. Dawers[1]
Department of Geology and Geophysics
The University of Edinburgh

Abstract

Previously published theory, which extrapolates fault and fracture population statistics observed in a one dimensional sample to two and three dimensional populations, is found to be of limited value in practical applications. We demonstrate here how significant the discrepancies may be and how they arise. There are two main sources for the discrepancies: (1) deviations from ideal spatial uniformity (spatial Poisson process) of a fault or fracture pattern, and (2) non-power law scaling of the size frequency distributions of the population. We show that even small fluctuations in spatial density, combined with variance in the estimator of population statistics, can lead to considerable deviations from the theoretical predictions. Ambiguity about power law scaling or otherwise of the underlying population is a typical characteristic of natural data sets and we demonstrate how this can affect the extrapolation of one-dimensional data to higher dimensions. In addition, we present new theoretical approaches to the problem of extrapolation when clustering of faults and fractures is explicitly considered. Clustering is commonly observed in the field as en echelon arrays of fault or fracture segments and we show how this property of natural patterns can be quantified and included in the theory. These results are relevant to building more realistic three dimensional models of the physical properties of fractured rocks, such as fracture permeability and seismic anisotropy.

[1]Now at: Department of Geology, Tulane University, New Orleans, LA 70118, USA

1 Introduction

It has previously been recognized that the observed size-frequency distribution of natural fault and fracture patterns depends on the sampling method used (e.g., Marrett & Allmendinger, 1991). For example, if the underlying population forms a three dimensional structural network, but only a two dimensional sample (e.g., a map) is analyzed, then many faults and fractures will be unrepresented in the resulting size-frequency distribution. Similarly, if the sample comes from a one dimensional traverse (e.g., a borehole) then under-sampling will be even more pronounced. It is only when the dimension of the sample equals the dimension of the underlying fault or fracture pattern that a true estimate of the distribution can be determined.

A standard way of making a correction for under-sampled distributions is to assume that faults and fractures are independently, uniformly distributed in space with a constant spatial intensity (Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998). This means that the probability of sampling a fault or fracture depends only on its size (e.g., length) and not on its spatial position. Moreover it implies that the size of a particular structure and its position relative to other structures within the overall pattern are unrelated. There are many field observations that indicate that these assumptions are unrealistic (e.g., Bour & Davy, 1999), as we review in more detail below.

The reason that the assumption of zero spatial correlation has proved so persistent in the literature is that it leads to a very simple result: for a self-similar fault or fracture population following a power law size-frequency distribution, as is commonly observed, the difference between the true power law exponent and that of a sample is an integer (Marrett & Allmendinger, 1991). The value of the integer is given by the dimension of the underlying fault network minus the dimension of the sample that is considered. Thus the integer takes the value zero if the pattern is fully sampled; the maximum integer value is 2.0, which corresponds to a 1D sample (i.e., a traverse) of a 3D pattern. In practical applications this simple result is potentially very useful: for example, borehole data on fault or fracture density may be extrapolated to construct a 3D structural network within a rock volume for which bulk properties such as permeability, brittle strain, or seismic velocity and anisotropy are required (Marrett, 1996; Berkowitz & Adler, 1998).

To our knowledge Malinverno (1997), in his work on the spatial distribution of turbidites, presents the only published attempt to improve or modify the assumptions on which the theory is based. In this paper we address this issue, building on the work of Malinverno (1997), but focusing specifically on sampling of fault and fracture populations. In the theoretical sections below we confine our analysis to faults in particular, using information on fault scaling properties from the literature. Faults are structural discontinuities across which there is significant shear displacement. However, in the discussion (Section 6) we show how our results are generally applicable to dilitant cracks, such as joints and veins, as well as to shear fractures.

The parameters which are usually used to describe the geometry of a fault population include: (1) fault size, which may be the length and /or the displacement across each fault,

(2) the orientation of each fault, and (3) the position of the center point (or barycenter) of each fault plane. It is these three properties, how they are related and how they scale within a particular fault population, that are the subject of this paper. Other information that may be used, concerns the nature of fault intersections, i.e., the cross-cutting relationships which may indicate the relative timing of formation of faults of differing orientation. The term fault population is used to describe a network of faults that is known to have developed within the same tectonic regime, and/or a single tectonic episode.

Previous authors have looked at the correlations within fault populations, for example between the spatial distribution of fault center points and the size-frequency distribution of fault lengths. Davy et al. (1992) and Bour & Davy (1999) have argued that the center points of natural fault patterns follow a fractal distribution and that there is a simple relationship between the fractal dimension of the spatial distribution and the power law exponent of the size-frequency distribution. We have found problems with applying their methodology to our data sets. The most significant problem is that an analysis of the fault center points alone suggests that the spatial distribution of some natural patterns is space filling (i.e., non-fractal) even though the faults are evidently arranged in en echelon arrays. En echelon geometry is a typical feature of natural fault and fracture systems and reflects the mechanics of growth of these structures. In an en echelon array the distance between neighboring fault tips is relatively small while the distance between the center points can be very large. Thus, a similar approach to that of Davy et al. (1992) could only be used to resolve en echelon alignments if additional information about the faults that make up the arrays is included, relating together the length, sub-parallel orientation and the relative position of faults in the arrays. The other problem with their method arises because the value of the scaling exponent of the size-frequency distribution, and indeed whether it is power law or not, can be difficult to determine unambiguously for many data sets.

In this paper we have developed a new approach for taking into consideration en echelon geometry as a specific example of spatial clustering, and have also explored the effects of non-power law scaling of the size-frequency distribution on sample statistics. We show that, although an integer difference between a sample and the underlying population distribution can be theoretically derived for a reasonable model (Marrett & Allmendinger, 1991), in practice the difference is non-integer. We find that significant deviations from the ideal case can be attributed to these two sources: i.e., a non-uniform spatial pattern (clustering) and/or a non-power law size-frequency distribution. The paper is designed with practical implementation in mind, in that we have chosen to address the issues that commonly confront a geologist/geophysicist when natural data sets are analyzed. Our results have important implications for extrapolating these data because of the power law scaling properties. This is because even moderate changes in the value of the power law exponent can result in order of magnitude variations in the numbers of fault or fractures which are predicted.

2 Uniform Spatial Distribution of Faults

A number of field studies of fault populations conclude that the size-frequency distribution of faults is best represented by a power law distribution (e.g., Special Issue of Journal of Structural Geology, Vol. 18, No. 2/3, 1996, and references therein). Let X represent the size of a fault, measured for example as the length L or the maximum displacement $D_{\max}$. The probability density function (pdf) of a power law (or fractal) distribution is given as the pdf of a Pareto distribution (Johnson et al., 1994)

$$f(x) = \frac{\beta x_0^\beta}{x^{\beta+1}}, \qquad x \geq x_0,\ x_0, \beta > 0 \tag{1}$$

where x_0 is the minimum fault size. The size-frequency distribution of a fault population is characterized by the power law exponent β, and this exponent can be estimated based on observed fault patterns. Faults are geometric objects embedded in three dimensions. However, observations are usually obtained from a lower dimension. For fault sizes observed in different dimensions, the value of the power law exponent β differs. Let C_1, C_2 and C_3 denote the values of the power law exponent β of the length distributions obtained in one, two and three dimensions respectively. Observations of faults in one dimension generally consist of single displacement measurements with no measurable length. In this case the length distribution is obtained from the displacement distribution of the faults being intersected by the traverse, as displacement and length are related (Watterson, 1986). Due to the under-representation of smaller faults in lower dimensions, the exponents differ according to $C_1 < C_2 < C_3$.

Fault patterns can be represented as marked point processes (Stoyan et al., 1995) where the fault positions form the point process and the marks are given as the fault sizes. In a Poisson point process with independent marks the fault positions are independently, uniformly distributed in space and fault size is independent of position. Under this assumption the following relationships between power law exponents in one, two and three dimensions are found:

$$\begin{aligned} C_2 &= C_3 - 1 \\ C_1 &= C_3 - 2 = C_2 - 1 \end{aligned} \tag{2}$$

(Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998, see also Appendix A, where the relationship between 1D and 2D is derived for a generalized size frequency distribution.). If C_1 or C_2 is estimated from observations, the power law exponent in a higher dimension is easily obtained by inversion of Eq. (2) as long as the assumption of spatial uniformity holds. If a relationship $D_{\max} \propto L^n$ is assumed between length and maximum displacement of faults (Cowie & Scholz, 1992a, b), the power law exponents of fault displacements are given as $C_i' = C_i/n$. The relationships between the power law exponents C_1', C_2' and C_3' for displacements in one, two and three dimensions

are then given as

$$C_2' = C_3' - \frac{1}{n}$$
$$C_1' = C_3' - \frac{2}{n} = C_2' - \frac{1}{n}. \quad (3)$$

In uniform rock types and tectonic settings the relationship between L and $D_{\max}$ is found to be linear; $n = 1$ (Cowie & Scholz, 1992a, b; Dawers et al., 1993). Schlische et al. (1996) find that the "global" dataset is also best fit by $n = 1$, although some authors earlier argued that $n > 1$ (e.g., Walsh & Watterson, 1988). In the case $n = 1$ the power law exponents of the displacement distributions equal the exponents of the length distributions, and Eq. (3) coincides with Eq. (2). For a general n, we use the relationship $D_{\max} \propto L^n$ to derive length distributions from displacements measured along 1D traverses.

Using Eq. (2) or (3) to extrapolate from one to two dimensions, or from two to three, can give unreliable results of the power law exponent in higher dimensions. Obviously, if the assumptions leading to the results are not fulfilled then the theoretical relationships will change. However, when power law exponents are estimated, the uncertainty lies not only in the fulfillment of the requirements behind Eq. (2) and (3), but also in the validity of the estimate. The variance of the estimator is a source of uncertainty, regardless of the spatial distribution of fault positions. This is illustrated through a simulation example. A 2D fault pattern consisting of 25,000 faults is simulated. A small section of the fault pattern is given in Figure 1a. The fault pattern is generated by first drawing fault center points from a uniform distribution on a two dimensional square of size $250,000\,m^2$. Next, orientations are drawn uniformly on the interval $[0°, 360°]$ and fault lengths are drawn from a power law distribution with exponent $C_2 = 1.6$ and minimum fault size $x_0 = 1m$. Although the pattern does not appear comparable to many field examples, it is typical of patterns generated under the assumption of uniform spatial distribution of fault locations. To study the length distribution obtained in 1D, data is collected along 1000 horizontal traverses at randomly chosen positions. For each traverse the power law exponent of the length distribution is estimated for those fault traces that are intersected, using a maximum likelihood estimator. The maximum likelihood estimator is known to be biased, even though it is commonly used (Clark et al., 1999). Figure 1b shows a histogram of the estimated exponents C_1.

The assumptions behind the result in Eq. (2) are fulfilled, and the theoretically predicted value of the power law exponent for lengths observed along the traverses is $C_1 = 0.6$. The estimated exponents are centered around the theoretical value, although slightly biased with an average of 0.62. However, the important point shown in Fig. 1b is that the standard deviation is 0.07 and that the bulk of the estimates lies in the range 0.5–0.75, giving a difference between C_2 (=1.6) and C_1 of 0.85–1.1. Thus in these examples the natural variations in the estimator can give a deviation of 15–25% between the estimated exponent and the theoretically predicted value. Repeated sampling studies give similar results, and similar conclusions are also reached if other fault orientations are considered, from uniformly distributed orientations over the interval $[0°, 360°]$ to all faults being paral-

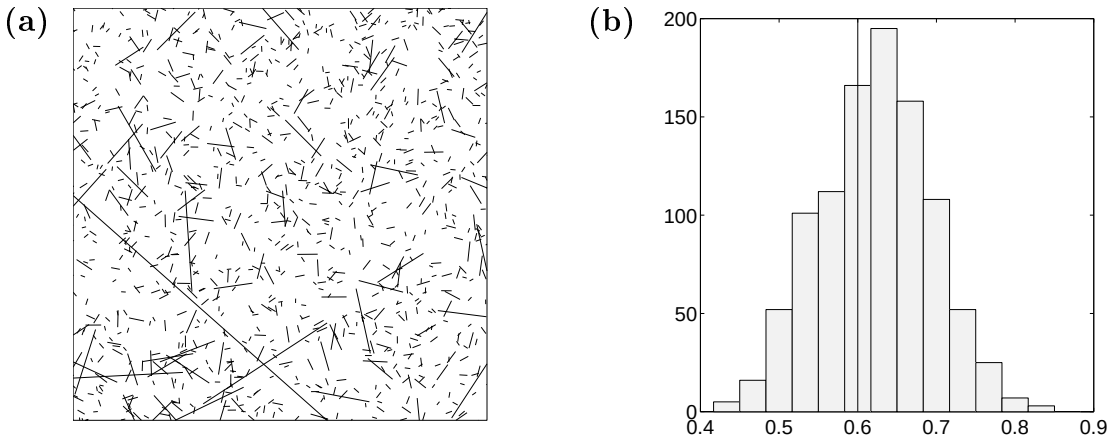


Figure 1: (a) Simulated fault pattern with a uniform spatial distribution and a power law length distribution with exponent $C_2 = 1.6$. (b) Histogram of estimated exponents C_1 from 1000 traverses. The solid line marks the theoretical value $C_1 = C_2 - 1 = 0.6$.

lel. The key issue here is the number of faults intersected by the traverse. In the simulation example, the number of faults intersected by the traverse lines varies considerably from 60 to 105 with a mean of 82. For traverse datasets containing a larger number of faults the bias and standard deviation will be lower, while for a decreased number of observations the uncertainty increases.

Clark et al. (1999) compare estimators for the power law exponents, and discuss inefficiencies in the geoscientists' traditional estimator. The standard way of estimating the power law exponent is by fitting a straight line to the plot of $\log N(x)$ versus $\log x$, where $N(x)$ is the number of faults larger than x. The estimates presented in Fig. 1b are obtained using a maximum likelihood estimator, which has known mean and variance (Clark et al., 1999). For $C_1 = 0.6$ and a number of observations ranging from 60 to 100, the standard deviation of the maximum likelihood estimator ranges from 0.08 to 0.06, in agreement with the empirical standard deviation of 0.07, found above. If the number of observations was increased to ~ 1000, the standard deviation of the maximum likelihood estimator would decrease to 0.02, reducing the error in the estimated power law exponent to $\leq 5\%$. But such large samples are unusual in natural data sets; samples sizes of tens to hundreds are much more typical.

3 Deviations from a Uniform Spatial Distribution

Two main assumptions behind the results in Eq. (2) and (3) are uniform spatial distribution of fault positions and power law size-frequency distribution. In this section the effect of deviations from a uniform spatial distribution is studied. Sub-parallel faults are considered, sampled along a traverse oriented perpendicular to the direction of fault strike (Fig. 2). The relationship $C_1' = C_2' - 1/n$ observed in a traverse through a two dimensional

fault pattern should be independent of the position of the traverse if the assumption of independently, uniformly distributed fault positions holds. Deviations from the uniform spatial distribution parallel to fault strike will alter the relationship between C_1' and C_2'. The spatial distribution is assumed to be uniform perpendicular to strike, i.e., along the the direction of the traverse, thus the length of the traverse is of no importance.

We use a similar approach to Malinverno (1997), who studied the statistical distribution of turbidite beds, and the relationship between the bed thickness distribution observed along a 1D traverse (a borehole) and the overall bed extent in 3D. Malinverno (1997) idealized turbidite beds as conical 3D bodies with a maximum thickness in the center, decreasing to zero thickness laterally. In this work we assume that turbidite bed thickness is analogous to variations in displacement over a fault plane, where the displacement typically is decreasing from a maximum near the center to zero at the fault boundaries (Walsh & Watterson, 1988). Malinverno (1997) considers vertical boreholes through horizontal sedimentary layers, while the fault data studied in this work are 1D transects across a map (Fig. 2). The geometric representation of turbidite bed thickness used by Malinverno (1997) is used here to represent the displacement profile along a fault trace in 2D. When the traverse intersects a fault, it intersects at an arbitrary distance r from the center point. The displacement observed in the traverse is then given by $d(r; D_{\max})$, where $D_{\max}$ denotes the maximum displacement. We can write the relationship between length and maximum displacement as $D_{\max} = \gamma L^n$. The profile is then given as (Malinverno, 1997)

$$d(r; D_{\max}) = D_{\max} \left\{ 1 - \left(2r \left(\frac{\gamma}{D_{\max}} \right)^{1/n} \right)^{\alpha} \right\}, \qquad 0 \leq r \leq \frac{1}{2} \left(\frac{D_{\max}}{\gamma} \right)^{1/n} \tag{4}$$

where the parameter $\alpha > 0$ controls the shape of the displacement profile. For $\alpha < 1$ the displacement profile is peaked at the center, for $\alpha = 1$ the profile is triangular while for $\alpha > 1$ the displacement increases rapidly near the end points and flattens out in the center part; see also (Malinverno, 1997). As $\alpha \to \infty$ the profile becomes rectangular. The displacement is assumed to be at its maximum at the center point of the fault, and all faults are assumed to be approximately parallel with an orientation perpendicular to the traverse.

We now consider a non-uniform spatial distribution of faults, where the location of the traverse has influence on the observed power law exponent C_1'. Malinverno (1997) introduces a distribution $f(r) \propto r^{\delta - 1}$, $\delta > 0$, for the distance from a sampling line to the center point of a turbidite bed. A similar distribution can be used for the distance from the traverse to a fault center point, measured perpendicular to the traverse. If $\delta = 1$ the fault positions are independently, uniformly distributed. If $\delta < 1$ the spatial distribution relative to the traverse deviates from the uniform distribution, and the faults are clustered in a self-similar fashion in the direction of the strike. The distribution $f(r)$ represents a decreased intensity of fault positions away from the traverse when $\delta < 1$, which can be interpreted as the traverse being located in a high intensity area of faults compared to the overall fault density. Reformulating the results of Malinverno (1997), the power law

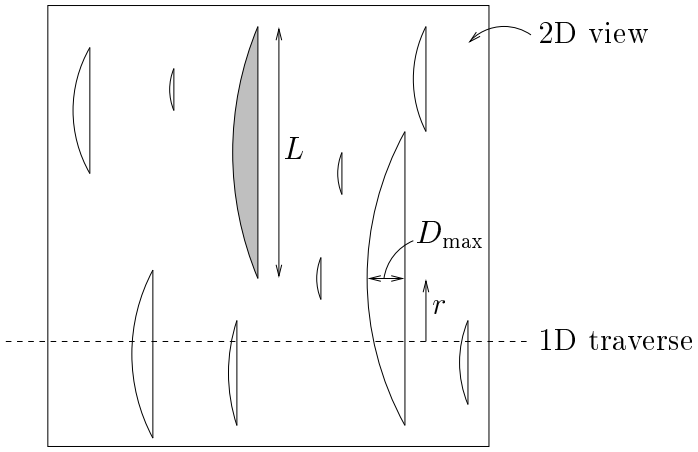


Figure 2: Illustration of 1D traverse line through a 2D fault pattern. The shaded area indicates the fault displacement profile of a fault of length L, Eq. (4). The displacement is at a maximum $D_{\max}$ at the fault center point, and the distance from the traverse line to the center point of an intersected fault is r.

exponent of displacements observed along the traverse is found to be

$$C_1' = C_2' - \frac{\delta}{n}. \tag{5}$$

The corresponding result for power law exponents $C_i = nC_i'$ of fault lengths is

$$C_1 = C_2 - \delta. \tag{6}$$

This result shows that if the spatial distribution perpendicular to the traverse line is not uniform, but scales in a self-similar fashion, the difference between power law exponents in two and one dimension differs from the theoretically obtained value in Eq. (2) and (3). For a non-uniform spatial distribution with $\delta < 1$, the difference between C_2 and C_1 will be less than the expected value 1. For $\delta = 1$ Eq. (6) and (5) reduces to Eq. (2) and (3) as expected. The relationships in Eq. (6) and (5) are independent of the shape parameter α, thus the choice of α is not critical, although if $n = 1$ then the fault population is self-similar with a constant value of α. For interacting and linked faults α is no longer a constant (Willemse et al., 1996), but the effect of this variation may be of secondary importance.

4 Deviations From Power Law Size-Frequency Distribution

We now examine the second major assumption that underlies Eq. (2) and (3), that of the power law size-frequency distribution. Many field data sets, in continental regions at least, have been interpreted as showing power law size scaling (see Special Issue of Journal of

Structural Geology, Vol. 18, No. 2/3, 1996). However, any natural data set is affected to a greater or lesser degree by the effects of limited resolution and censoring. The former leads to under-sampling of small faults while the latter leads to inadequate sampling of faults that are comparable in size to the dimensions of the sampling domain (e.g., a map or cross-section). Although various corrections have been proposed to minimize and quantify these effects (e.g., Pickering et al., 1995, 1996), the corrections assume that the true distribution is power law. However, some fault populations, such as those formed at mid-ocean ridges (Cowie et al., 1994) are better described by an exponential distribution and Ackermann et al. (1997) and Spyropoulos et al. (1999) have documented both power law and exponential distributions in analog experiments of fault growth (see also Ackermann, Withjack & Schlische, The effects of mechanical layer thickness on the systematics of an evolving population of normal faults in experimental models, manuscript submitted to J. Struct. Geol., September 1999).

A general case of deviations from a power law distribution of fault size will not be considered, but the special case of fitting a power law exponent to data that is in fact exponentially distributed is discussed, as this is a common ambiguity in interpreting fault population data (Ackermann et al., 1997; Cowie, 1998). The relationship between the length distributions in two dimensions and along a 1D traverse is studied, to quantify the under-representation of small faults for exponentially distributed data. Note that if fault displacements were to be analyzed instead, the displacement profile must also be included in the calculations, giving analytically non-tractable expressions.

Let $N(x)$ be the cumulative number of observed faults with length larger than x. If the fault length follows a power law distribution, Eq. (1), a plot of $\log N(x)$ versus $\log x$ gives a straight line with slope $-\beta$. If the fault length follows an exponential distribution with intensity $\lambda = (\text{mean of } x)^{-1} = < x >^{-1}$ and probability density function

$$f_E(x) = \lambda \exp(-\lambda x), \qquad x \geq 0 \tag{7}$$

the log-log plot produces a concave shape to the curve. However, a similar type of concave curve can be the result of plotting poorly sampled data from a power law distribution, where truncation, censoring and finite range effects have led to a flattening of the curve for small x and a steep fall off for large x (Pickering et al., 1995). Thus data from a truly exponential distribution can mistakenly be assumed to follow a power law, using the center part of the log-log plot to fit a power law exponent to the data. Note that the value of λ depends on the scale of the observations (in contrast to the power law exponent).

Suppose the fault lengths in two dimensions follow an exponential distribution, but a power law exponent C_2 is fitted to the data. This exponent is then compared to a power law exponent C_1 fitted to data along a one dimensional traverse through the observation area. Since a plot of $\log N(x)$ versus $\log x$ for exponentially distributed data does not give a straight line, the slope of a local line-fit will vary depending on x. When a power law exponent is calculated, a straight line is fitted to the center part of the data, assuming this part to be approximately straight. The fitted power law exponent will be close to the absolute value of the slope of the true curve for this part of the data. At length x the

theoretical absolute value of the slope is

$$C_2(x) = \lambda x. \tag{8}$$

If the fault length in two dimensions follows an exponential distribution, Eq. (7), the length of faults intersected by a one dimensional traverse follows a gamma distribution (Johnson et al., 1994) with pdf

$$f_G(x) = \lambda^2 x \exp(-\lambda x), \qquad x \geq 0, \tag{9}$$

see Appendix A. The absolute value of the slope of the curve $\log N(x)$ versus $\log x$ at length x is

$$C_1(x) = \frac{(\lambda x)^2}{\lambda x + 1}, \tag{10}$$

as shown Appendix B. The resulting difference in exponents for the 2D and 1D population at length x is then

$$C_2(x) - C_1(x) = \frac{\lambda x}{\lambda x + 1} < 1, \tag{11}$$

independent of the value of λ. One of the assumptions leading to the result in Eq. (2) is no longer valid and consequently $C_2 - C_1$ differs from 1. We can demonstrate how important the difference may be by considering what happens when these power law exponents are derived by fitting straight lines at the mean values $< x >= 1/\lambda$ (in 2D) or $< x >= 2/\lambda$ (in 1D) in the exponential distribution in 2D, Eq. (7), and the corresponding distribution in 1D, Eq. (9), respectively. These choices of x give $C_2(x) - C_1(x)$ values of 1/2 and 2/3 respectively, regardless of the value of λ. Thus for typical values of fault lengths of the given distributions, the difference in slopes between one and two dimensions is significantly smaller than 1.

Ortega & Marrett (2000) show a good example of this situation in their analysis of natural fracture patterns in sandstones. Their Figure 13 compares the size frequency distribution of a two dimensional fracture population with that of a one dimensional sample of the same population. They fit both distributions with a power law even though they show that an exponential fit to the 2D data is equally good. The power law exponents that they calculate for the distributions are 0.98 and 1.25 in 1D and 2D respectively. Thus, in this real example the difference $C_2 - C_1$ is only 0.28.

Based on the result in Eq. (11) it can also be concluded that if the distribution of fault length is neither power law nor exponential, but a mixture of the two distributions (Ackermann et al., 1997; Spyropoulos et al., 1999), the difference in fitted exponents in two and one dimension will be less than 1.

5 Interacting and Linked Fault Structures

In two dimensions faults are observed either as relatively isolated structures, or as segments forming en echelon arrays. Gupta & Scholz (2000) show how the displacement profile

along a segment becomes gradually distorted compared to that of an isolated fault as the ratio of overlap (O) to separation (S) between fault tips increases and the degree of elastic interaction increases (see also Peacock & Sanderson, 1991; Willemse et al., 1996). Eventually the ratio O/S is limited by the inability of segments to grow into regions of stress reduction around the interacting segments and the highly distorted displacement profiles, when summed together, approach that of a relatively isolated fault (Willemse et al., 1996; Gupta & Scholz, 2000). En echelon arrays are thus easily detected in 2D observations, partly because of profile shape of individual segments, but also because some overlapping segments may be physically linked in places by smaller connecting faults and fractures. When faults are observed along a 1D traverse however, it is more difficult to determine whether two structures occurring close in space are two isolated faults or are segments of the same fault structure along strike. Wojtal (1996) showed how criteria used for defining linked fault segments in a 1D data set of thrust faults from the Cumberland Plateau in Tennessee, has an important influence on the estimated power law exponent in an analysis of fault size-frequency distributions.

Figure 3 illustrates the ambiguity in the interpretation of fault linkage on a 1D traverse. Fig. 3a shows the throw profile along a fault array modified from Dawers & Anders (1995), and Fig. 3b gives a map view of the fault array showing the location of two traverses $x-x'$ and $y-y'$ where the total throw is comparable. Fig. 3c compares the two 1D traverses across the fault. Along $x-x'$ four separate structures are seen across a zone $\sim 500m$ wide, whereas along $y-y'$ a single fault plane is observed. Without 2D information it is not possible to judge from traverse $x-x'$ alone whether the four separate faults are linked along strike, even though it is clear from the summed displacement profile that they do form part of a single larger structure.

One of the main assumptions behind the results in Eq. (2) and (3) is not satisfied for fault segments in en echelon arrays, since these represent a type of clustering of fault segments. In this section we consider a fault pattern that consists of a mixture of fault arrays and isolated faults. For simplicity it is assumed that the results in Eq. (2) and (3) hold for isolated faults or fault arrays separately, and also for the individual segments of the arrays, but not for the overall pattern.

An en echelon array is composed of a number of shorter fault segments. We separately define the lengths of the individual segments and the length of each fault array. If the length distribution of arrays is compared to the length distribution of the individual fault segments, it is obvious that the length distribution of the segments will have a greater proportion of small faults than that of the arrays. Suppose both length distributions are power law distributions, with exponent β_S for the segments and β_A for the arrays. Since the value of the power law exponent decreases as the proportion of large faults increases, this implies that $\beta_S > \beta_A$. The same argument applies for the displacement. The displacement of a fault array is a sum of displacements on the smaller fault segments, thus the linked structures tend to have larger displacements than the shorter segments of which the array is composed (see Fig. 3).

In this work we distinguish between isolated faults, fault arrays and fault segments which comprise the arrays. As we have already discussed, the segments making up an

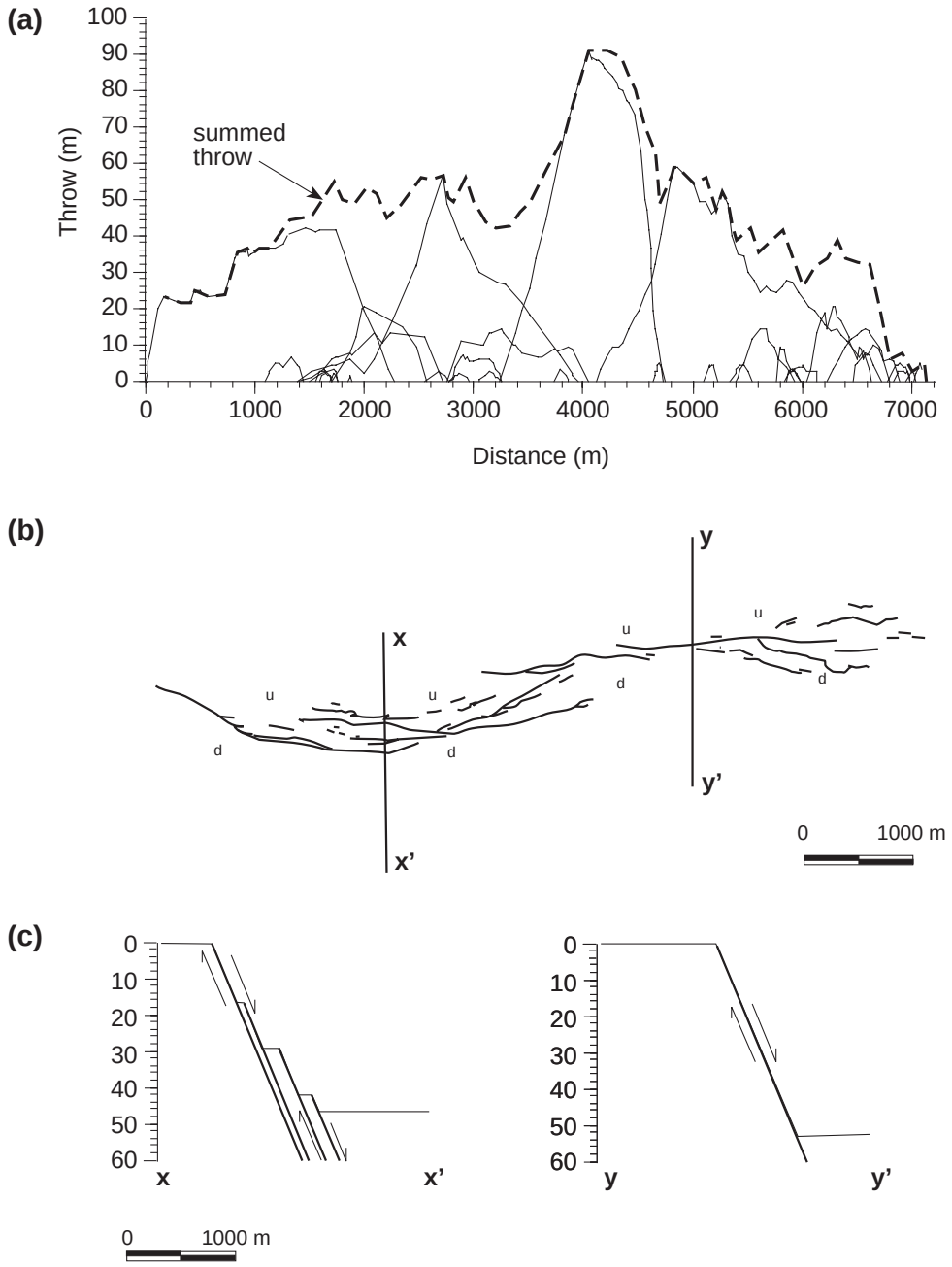


Figure 3: (a) Throw profile along a linking fault array from the Volcanic Tableland in eastern California modified from Dawers & Anders (1995). Solid lines show throw variations along individual segments of the array, dashed line shows the total throw along the entire array obtained by summing together the throws of the separate segments. (b) Map view of fault array showing the location of two traverses $x - x'$ and $y - y'$ shown in (c). (c) Comparison of the two 1D traverses $x - x'$ and $y - y'$ across the fault where the total throw is comparable. See text for further discussion.

array may have displacement profiles differing from the typical profile of an isolated fault, while the shape of the overall (summed) displacement profile of the array is comparable to that of isolated faults (e.g. Peacock & Sanderson, 1991; Dawers & Anders, 1995; Willemse et al., 1996). For simplicity it is assumed in this work that isolated faults and segments comprising arrays have the same size-frequency distribution, whether we compare length or displacement. Consequently we are also making the assumption that the maximum size of the isolated faults are smaller than that of the fault arrays, and that the populations of isolated faults and segments that comprise arrays are comparable.

Let X represent the maximum displacement of a fault observed in two dimensions, and suppose X follows a power law distribution. The power law exponent depends on whether X represents an isolated fault or an en echelon array. If X represents an isolated fault the power law exponent is β_S, and the pdf of X is denoted $f(x;\beta_S)$. If X represents an array, the power law exponent is β_A and the pdf is $f(x;\beta_A)$. The overall size-frequency distribution of fault displacements observed in two dimensions is a combination of the two distributions:

$$f(x) = (1-\theta)f(x;\beta_S) + \theta f(x;\beta_A) = (1-\theta)\cdot\frac{\beta_S x_0^{\beta_S}}{x^{\beta_S+1}} + \theta\cdot\frac{\beta_A x_0^{\beta_A}}{x^{\beta_A+1}}, \qquad 0\le\theta\le 1 \tag{12}$$

where θ is the proportion of arrays in the fault pattern, given as the number of arrays divided by the total number of faults. The parameter θ is determined by imposing a criterion on the 2D fault observations which separates isolated structures from those that are interacting and/or linked segments within an array. Here we use a simple criterion for classifying two faults as strongly interacting, and hence effectively linked, based only on the value of the O/S ratio (Fig. 4). The observed values of O/S of fault arrays typically lie in the interval 1-10 (Aydin & Schultz, 1990; Huggins et al., 1995; Gupta & Scholz, 2000; Schultz, 2000). A more rigorous criterion would include information on the relative sizes of the segments (e.g., Gupta & Scholz, 2000; Schultz, 2000), but for the purposes of this paper the O/S ratio provides a convenient first approach to this problem.

Let N_I be the number of isolated faults at a given data resolution, and let N_A be the number of fault arrays. The value of θ is then given as

$$\theta = \frac{N_A}{N_I + N_A}. \tag{13}$$

Note that the value of θ will depend on the linkage criterion used to interpret the fault pattern, and also on the resolution of the data because both of these factors will influence the ratio calculated in Eq. (13). Figure 5 gives an illustration of three fault patterns with

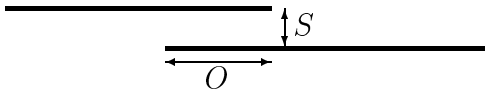


Figure 4: Definition of overlap O and separation S.

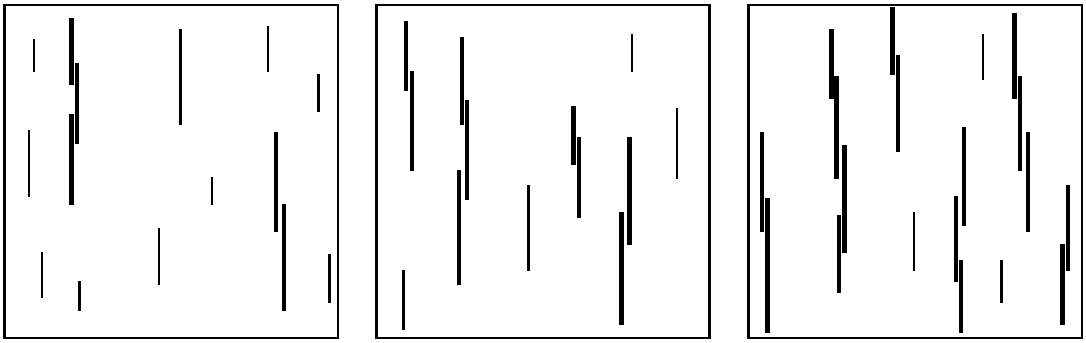


Figure 5: Three fault patterns with different proportion θ of linked fault structures. The thick solid lines represent linked faults with O/S ratios in the range 3–8.

a varying value of θ. For a normal fault population in the Volcanic Tableland of eastern California (Dawers et al., 1993; Dawers & Anders, 1995), the proportion θ is thought to be less than 0.5. For this field data, the values of O/S in our linkage criterion lie in the interval 3–10 (Dawers, 1996), as indicated by the asymmetry of the displacement profiles of strongly interacting faults and/or physically linked faults.

Figures 6, 7 and 8 give an illustration of the combined distribution function (12) and the appearance of data originating from this distribution. The observations in Figures 6 and 7 are generated using statistical sampling techniques. Figures 6 and 7 show log-log plots of relative cumulative number $N(x)/N_T$ versus size x, where N_T is the total number of faults. The expression $N(x)/N_T$ is an estimate of the cumulative distribution function $S(x) = \text{Prob}\{X \geq x\}$, see Appendix B. Figure 8 shows the cumulative distribution functions $S(x)$ for different values of θ.

The combined distribution is only a power law distribution in the extremes $\theta = 0$ or $\theta = 1$, where it equals one of the two separate distributions $f(x;\beta_S)$ or $f(x;\beta_A)$. For intermediate values, $0 < \theta < 1$, Eq. (12) does *not* give a new power law distribution. Thus a plot of $\log N(x)$ versus $\log x$ does not give a straight line according to the theory. However, as Figure 6 shows, data from the distribution $f(x)$ appear to be approximated quite well by a straight line on log-log plot. Figure 6 shows a log-log plot of an increasing number of observations generated from the distribution $f(x)$, using power law exponents $\beta_S = 1.8$ and $\beta_A = 1.3$ that are within the range typical for two dimensional fault patterns (e.g., Special Issue of Journal of Structural Geology, Vol. 18, No. 2/3, 1996, and references therein). If these were real observations, with the theoretical expression in Eq. (12) unknown, most geologists would conclude that the data originate from a power law distribution (Scholz et al., 1993).

Figure 7 compares the combined dataset in Fig. 6b with two datasets generated from the original power law distributions with exponents β_S and β_A. On a cumulative log-log plot, the combined distribution is located between the two distributions of which it is composed. Figure 8 shows a log-log plot of the theoretical cumulative distribution function

$S(x)$ for the combined distribution, together with the corresponding curves for $f(x;\beta_S)$ and $f(x;\beta_A)$, for an increasing value of the proportion θ. The curve of $\log S(x)$ for the combined distribution is slightly convex. If however the data are assumed to follow a power law distribution, it is observed from Figs. 7 and 8 that an apparent power law exponent β_C of the combined distribution $f(x)$ lies between β_A and β_S, thus $\beta_A < \beta_C < \beta_S$. The estimated exponents of the observations in Fig. 6 are plotted as straight lines, together with the curved theoretical line. The curve of $\log S(x)$ in Fig 8b, using $\theta = 0.25$, corresponds to the samples shown in Figs. 6 and 7. As a result of the slightly convex form of the curve given by Eq. (12), the straight line fitted to the observations fall below the theoretical line as the fault size increases (Fig. 6). When the sample size increases, the fitted power law tends to under-estimate the number of large faults (Fig. 6d).

Table 1 gives estimated power law exponents β_C for an increasing number of observations and a varying proportion θ of fault arrays. Each estimate is based on 10,000 samples generated from the combined distribution (12), using $\beta_S = 1.8$ and $\beta_A = 1.3$, and the mean and standard deviation of the estimates are listed in the table. It is observed from the table that β_C approaches β_S for small ratios θ, and approaches β_A as θ increases, in agreement with the results shown in Figure 8. It is also observed from Table 1 that as the number of observations decreases, the estimated power law exponent β_C increases. This is due to the fact that the theoretical line of $\log N(x)$ is steepest for small fault sizes (Fig. 8). This region of the curve dominates the small datasets, which do not span more than one order of magnitude (Fig. 6). Figure 9 summarizes the results from Table 1, for observations consisting of 500 faults. The error bars show the uncertainty in the estimate of the power law exponent β_C. A larger number of observations would produce similar plots, only with shorter error bars, while a smaller number of observations would give a wider range of uncertainty. Note that the exponent β_C appears to decrease approximately linearly with θ, but the relationship is not exactly linear. Our theoretical work shows that, in general, the relationship between β_C and θ is not expected to be linear.

As pointed out above, the distribution $f(x)$ in Eq. (12) is not strictly a power law distribution. Assume however that the deviations from a power law is small enough to give

Sample	Ratio θ of linked fault structures, Eq. (13)				
size	0.10	0.25	0.50	0.75	0.90
100	1.751 (0.180)	1.657 (0.173)	1.525 (0.160)	1.412 (0.146)	1.354 (0.139)
500	1.736 (0.078)	1.646 (0.075)	1.513 (0.070)	1.399 (0.063)	1.340 (0.061)
1000	1.734 (0.055)	1.643 (0.054)	1.511 (0.049)	1.399 (0.045)	1.339 (0.043)
10,000	1.734 (0.018)	1.642 (0.017)	1.510 (0.016)	1.397 (0.014)	1.337 (0.014)

Table 1: Estimated power law exponents β_C based on 10,000 generated samples of increasing sample size from the combined distribution in Eq. (12). Different values of the proportion θ of linked fault structures are used. The table gives the mean value and standard deviation of the estimates, calculated from the 10,000 samples. The power law exponents $\beta_S = 1.8$ and $\beta_A = 1.3$ are used for isolated faults and fault arrays.

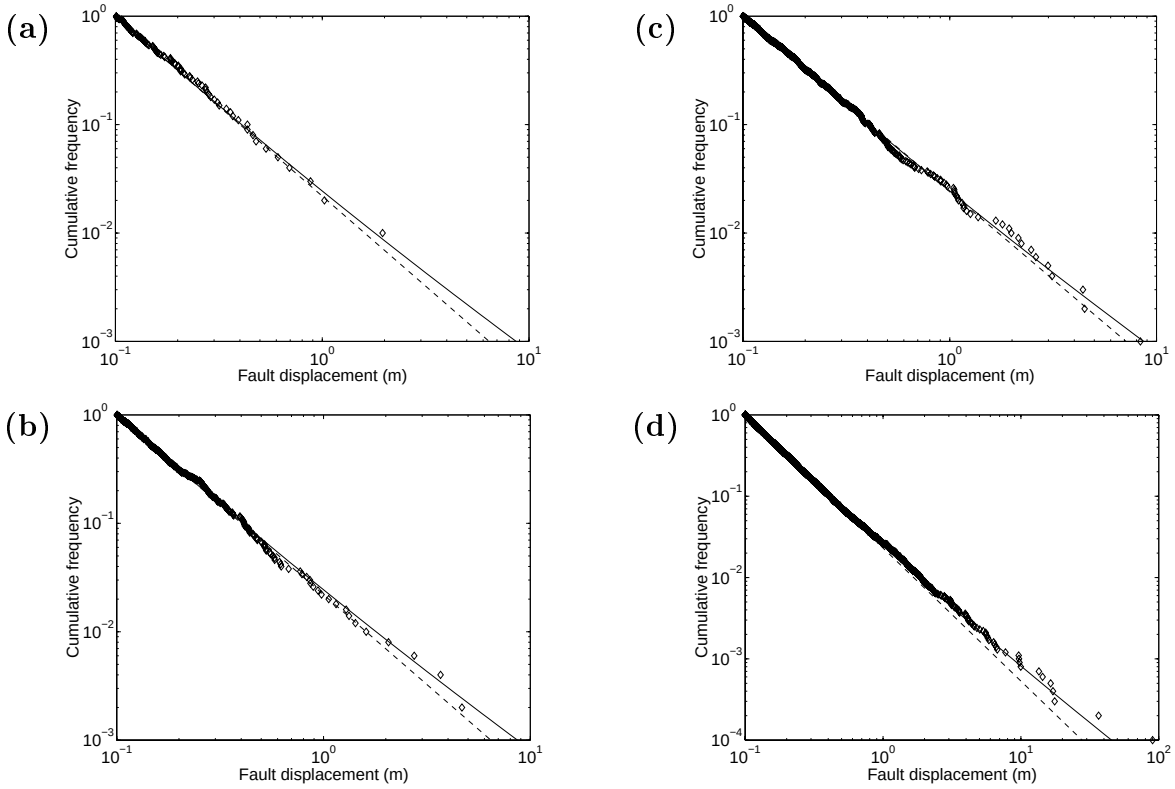


Figure 6: Samples generated from the combined distribution $f(x)$ in Eq. (12), with $\beta_S = 1.8$, $\beta_A = 1.3$ and $\theta = 0.25$. The number of observations in each sample is (a) 100, (b) 500, (c) 1000 and (d) 10,000 and the estimated power law exponents are (a) $\beta_C = 1.66$, (b) $\beta_C = 1.66$, (c) $\beta_C = 1.62$ and (d) $\beta_C = 1.64$. The estimated exponents are plotted as straight dashed lines, and the theoretical lines, obtained from Eq. 12, are plotted as solid lines.

$C_1' \approx C_2' - 1/n$ for the apparent power law exponents. The fault pattern in 2D, consisting of a mixture of isolated faults and en echelon arrays, has an apparent power law exponent of β_C. The corresponding exponent for observations along a traverse should then be $\beta_C - 1/n$, according to Eq. (3). However, if no linkage criterion is used along the traverse, and all observations are treated as single faults, the predicted power law exponent is $\beta_S - 1/n$, since isolated faults and single fault segments in 2D have exponent β_S. Comparing the 2D data with exponent $C_2' = \beta_C$ and 1D data with exponent $C_1' = \beta_S - 1/n$ thus gives a difference of

$$C_2' - C_1' = \frac{1}{n} - (\beta_S - \beta_C) < \frac{1}{n} \tag{14}$$

since $\beta_S > \beta_C$. In Section 3 clustering along strike was considered, and a new theoretical expression for the difference $C_2' - C_1'$ was obtained. A comparable expression based on Eq. (14) can not be derived, since the exponent β_C is only an apparent exponent, not given by an analytical expression. An estimate of β_C will depend on the situation: on all

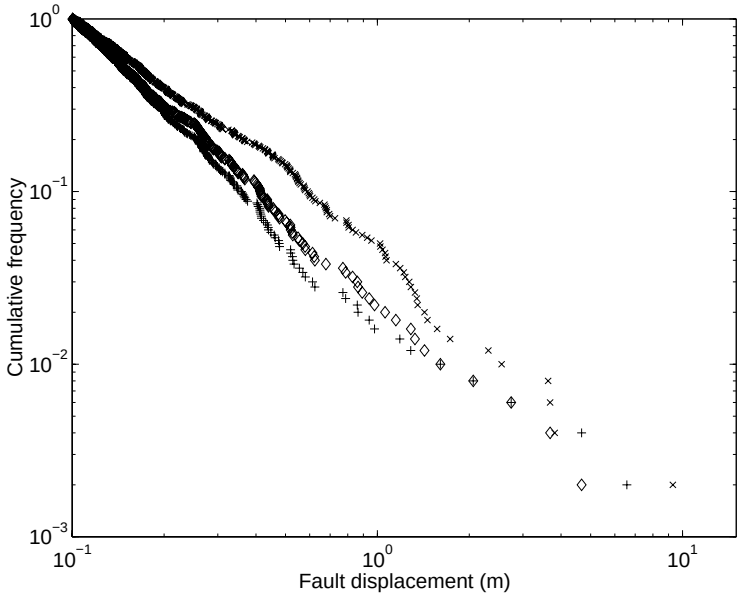


Figure 7: Comparison of the combined dataset in Fig. 6b (plotted as $\diamond$) with two datasets generated from the distributions $f(x;\beta_S)$ (plotted as $+$) and $f(x;\beta_A)$ (plotted as $\times$). The power law exponent of single faults is $\beta_S = 1.8$ and for fault arrays $\beta_A = 1.3$. The estimated power law exponent of the combined distribution is $\beta_C = 1.66$.

parameter values, the ratio θ and the number of observations (Tab. 1). Thus in order to estimate the difference between C_2' and C_1', an objectively measured 1D sample needs to be re-interpreted using some linkage criterion in an attempt to obtain 1D data corresponding to the distribution in Eq. (12).

There are two possible ways to do this: (1) by taking a purely statistical approach to linkage probability, as outlined in Appendix C, or (2) by re-evaluating the 1D data from known or potential linkages (e.g., Dawers, 1996; Wojtal, 1996; Schultz, 2000). Concerning the latter approach it is worth noting that for many published examples of 1D fault population data, the size-frequency distribution is obtained for a single stratigraphic horizon within a cross-section (e.g., Walsh et al., 1991, Marrett & Allmendinger, 1992). Thus, there is potentially some two dimensional information that may be extracted by examining the upper and lower tips of the faults which have been measured to assess how many are en echelon with other faults higher or lower within the section (e.g., Wojtal, 1996). Then using an overlap-separation criterion for linkage in the plane of the cross-section, and assuming that this applies in plan view, those faults belonging to arrays can be separated from those that are isolated to obtain a value for θ. The key assumption here is that evidence for dip-linkage in the cross-sectional view indicates potential linkage also occurring in plan view; a reasonable starting assumption but one which has not been independently verified.

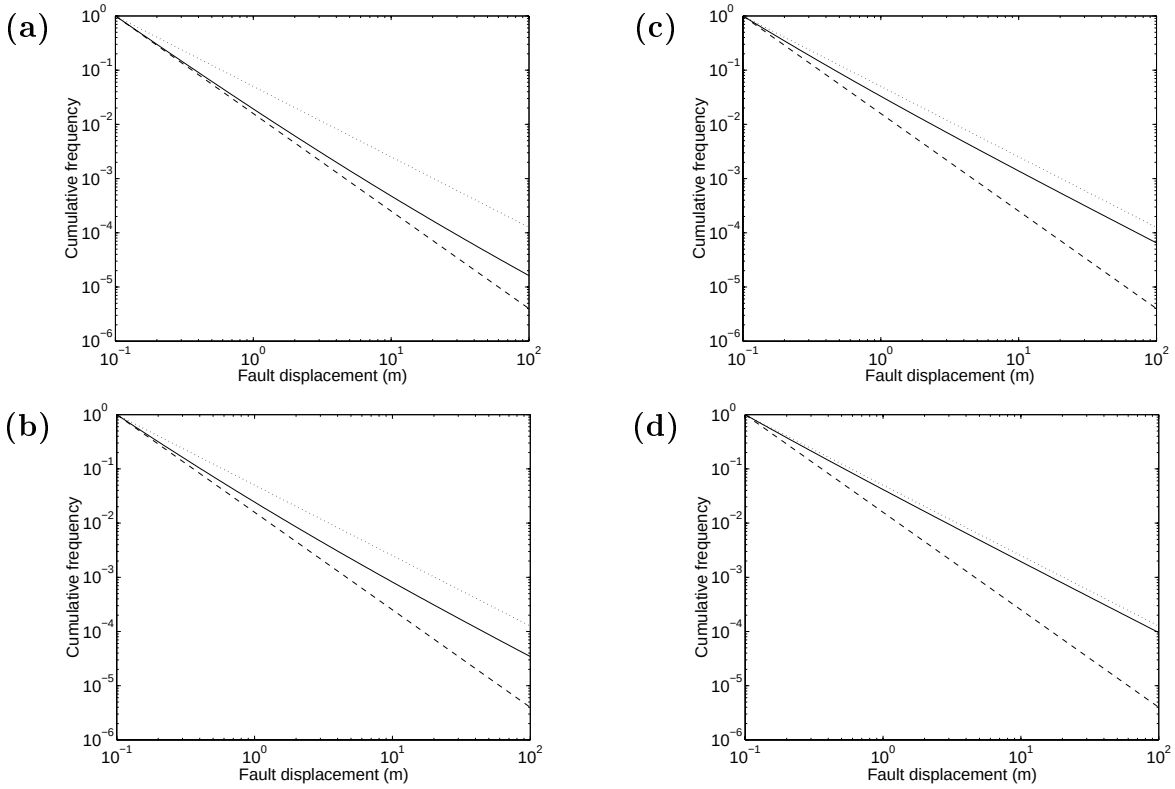


Figure 8: Log-log plot of cumulative distribution function $S(x)$ for the two distributions $f(x;\beta_S)$ (dashed lines) and $f(x;\beta_A)$ (dotted lines) with $\beta_S = 1.8$ and $\beta_A = 1.3$, and for the combined distribution $f(x)$ in Eq. (12) (solid lines). The lines are plotted using the ratios (a) $\theta = 0.1$, (b) $\theta = 0.25$, (c) $\theta = 0.5$ and (d) $\theta = 0.75$.

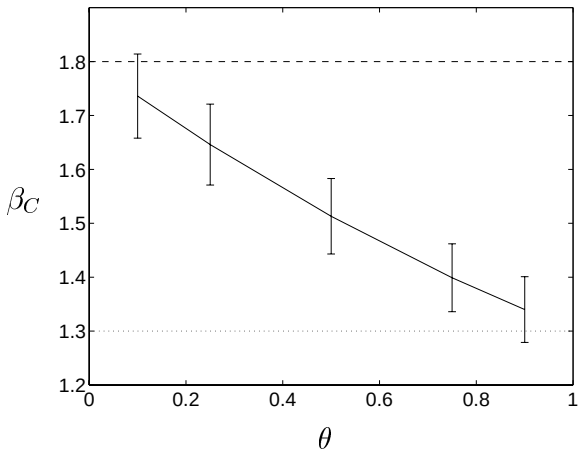


Figure 9: Variations in the estimated power law exponent β_C as a function of the ratio θ of linked fault structures (Tab. 1). The error bars show one standard deviation of the estimator for datasets of size 500 faults. The horizontal dashed line shows the value of the power law exponent $\beta_S = 1.8$, the dotted line the exponent $\beta_A = 1.3$.

6 Discussion

We have shown that there are several practical situations in which the addition or subtraction of an integer value from the power law scaling exponent does not provide an adequate representation of how fault populations in 1D, 2D and 3D are related. In the preceding theoretical sections we have been concerned with fault populations specifically. However, our results are generally applicable to joints, veins and shear fractures because these show similar scaling properties, and present similar difficulties in analysis and interpretation. Vermilye & Scholz (1995) show, for example, that tensile fractures and veins exhibit a similar relationship between aperture and length as we have used for fault displacement and length. Segall & Pollard (1983), in their analysis of joint patterns, found length-frequency distributions very similar to those found for fault populations. Their maps also clearly show the en echelon arrangement of joints, and they discuss some of the criteria for defining which joints are linked arrays (see also Ortega & Marrett, 2000).

The theory we have presented focuses on the extrapolation of 1D data to 2D and in particular the effects of spatial variations in fault density and en echelon clustering. These effects lead to a difference between C_2 and C_1 that is < 1.0. The implication of this is that if 1D data are extrapolated to 2D without considering clustering then the relative numbers of small scale faults and fractures are overestimated in the 2D distribution, the relative number of large scale structures is underestimated, and the size of the largest fault is also underestimated. Moreover, not only are the numbers of small fractures overestimated, the absence of spatial correlation information means that these small fractures are randomly positioned in the extrapolation to a 2D pattern, whereas in natural patterns they tend to be aligned in arrays.

Even though 2D data allows some evaluation of en echelon clustering and linkage, the same problems arise in going from 2D to 3D if the morphology of the fault surfaces and the network of faults are three dimensional. Thus the same effects will be seen as we have demonstrated here. Berkowitz & Adler (1998) generated 3D joint networks from 2D field observations using the established relationships given in Eq. (2). They then compared 2D slices through the model 3D volume with the original field maps, in terms of both the predicted 2D exponent and the visual appearance of the patterns. The original field measurements of joint length do take into consideration 2D linkage between closely spaced en echelon joints (Segall & Pollard, 1983). However, in the extrapolation to 3D the spatial locations of the joints are assumed to be independently, uniformly distributed, although this assumption is not tested (Berkowitz & Adler, 1998). For several cases the calculated exponents agree quite well, and discrepancies can largely be attributed to having small numbers of joints in the data sets which make the statistics less reliable. This agreement is expected since their 3D simulated joint patterns fulfill the requirements behind the result $C_3 - C_2 = 1$. However visual comparison of the real and modeled patterns shows two interesting features: Firstly, the real pattern exhibits zones of more concentrated jointing in the 2D plane separated (perpendicular to strike) by zones that are devoid of joints, and secondly, strike-parallel alignment and segmentation of the natural joints is very common. Neither of these properties is captured by the simulations. However, these are exactly the properties which will strongly influence the porosity structure, permeability and seismic velocity of fractured rock volumes.

Marrett (1996) discusses the impact of power law scaling of fault or fracture populations on "aggregate" properties, such as total fracture porosity, permeability, total fracture surface area, total fault strain, and shear-wave anisotropy. He concludes that, except for fracture surface area, larger fractures/faults contribute more to aggregate properties than do the smaller ones for most values of the scaling exponents. However, these conclusions depend on the value assumed for the power law exponents in 3D and on the assumption that the spatial pattern of the fracture or fault population originate from a uniform distribution. The present work has shown that the conclusions of Marrett (1996) may need to be re-evaluated in the light of spatial correlation information which has previously been ignored.

Hatton et al. (1993) derived an empirical relationship between 2D and 3D scaling exponents of the form $C_3 = a_1 C_2 + a_2$ where a_1 and a_2 are constants. According to their empirical analysis $a_1 = 1.3$ and $a_2 = -0.2$, whereas comparison with the present work would suggest $a_1 = 1.0$ and $0 \leq a_2 \leq 1.0$. Hatton et al. (1993) obtained this relationship from rock deformation experiments conducted under a variety of dry, damp and wet conditions which were designed to examine the differences in sub-critical crack growth processes in the presence/absence of a fluid. Estimates of C_3 were calculated from the size frequency distribution of acoustic emissions produced during the experiments whereas C_2 was obtained from populations of cracks observed in 2D slices of the deformed samples. Values of these exponents from both wet and dry experiments were combined together to obtain the relationship between C_2 and C_3, even though the morphology of the crack populations differed significantly. For the dry experiments the deformation was localized

onto longer more continuous cracks, whereas the wet experiments showed a distributed pattern of shorter cracks. Therefore, the values that they obtained for a_1 and a_2 may be due to combining the results of the different types of experiments. If instead the wet and dry experiments had been analyzed separately then, according to our theory, $a_2(\text{dry}) < a_2(\text{wet}) \leq 1.0$ due to the differing degree of localization in the two types of experiments, whereas a_1 should equal 1.0 for both.

7 Summary and Conclusions

The two key assumptions of the published theory (Heffer & Bevan, 1990; Marrett & Allmendinger, 1991; Piggott, 1997; Berkowitz & Adler, 1998), are that (1) the spatial distribution of faults and fractures can be modeled as a Poisson process where the locations of the individual structures are independently, uniformly distributed with a constant spatial intensity, and (2) that the size frequency distribution is a power law distribution. We find that, even when assumptions (1) and (2) are fulfilled, the difference of the estimated exponents from one and two dimensional samples of simulated fault patterns can deviate by 10–15% from the theoretically predicted difference for datasets consisting of 50–100 faults, using a maximum likelihood estimator for the exponent. For datasets consisting of $\sim$ 1000 faults, the error is reduced to $\leq 5\%$. This deviation is due to the variance in the estimator which is used to calculate the exponent, and due to local fluctuations in fault density which arise naturally even for a spatial pattern that is statistically uniform. We show further that ambiguity concerning assumption (2), due to the effects of limited resolution and censoring on natural fault patterns, can lead to even larger deviations from the results theoretically predicted by previous workers. Specifically, we have studied the effect of fitting a power law exponent to data originating from an exponential distribution, to illustrate how an incorrect assumption of power law scaling can lead to deviations from the results in Eq. (2) and (3). The deviation could be as large as 50%.

An important aspect of this paper concerns the effect of fault clustering on the relationship between fault populations sampled in one, two or three dimensions. In other words, we have studied in detail the impact of violating the assumption of spatial uniformity (assumption (1) above). In the first case we follow the approach of Malinverno (1997), and consider fault clustering perpendicular to a 1D traverse line which crosses a sub-parallel pattern of faults perpendicular to fault strike. If the distribution of fault centre points perpendicular to the traverse line is uniform, then $C_2 - C_1 = 1$ as previous theory predicts. However, when the fault centre points cluster in a self-similar fashion, the scaling parameter describing the spatial distribution modifies the relationship such that $C_2 - C_1 < 1$, depending on the degree of clustering. In the second case we consider the effect of fault clustering in a direction parallel to the 1D traverse line. This type of clustering is widely recognised in the field as en echelon arrays of fault segments which link as the faults grow during progressive deformation. We study the effect of en echelon clustering theoretically, and show how lack of information on linkage in 1D leads to problems when relating power law exponents obtained from 1D data to the corresponding exponent in 2D. The problem

of relating 1D power law exponents to 2D exponents is that without any linkage criterion in 1D all faults are assumed to be isolated, while in 2D parameters such as the ratio of fault tip separation to overlap are often used to link fault segments together into longer structures.

To show how this affects the statistics we represent the overall fault population as a combination of isolated faults and en echelon fault arrays. En echelon arrays are those comprised of segments with overlap-to-separation ratios between their tips that have values within a specified interval (Aydin & Schultz, 1990; Huggins et al., 1995; Gupta & Scholz, 2000). Separately, these two populations have different size-frequency distributions, with a higher proportion of small faults in the population of isolated faults. If the two separate populations are assumed to be power law distributions, represented by two different power law exponents, the combined distribution can be approximated by a third power law distribution. The combined distribution can be fitted by a power law exponent smaller than the exponent of the population of isolated faults, but larger than the exponent of the fault arrays. The relationship between these power law exponents is not found analytically, but is found via simulated fault patterns to depend on the ratio between the number of en echelon fault arrays and the total number of faults (θ), and thus on the linkage criterion and the resolution of the data.

The appearance of the combined distribution is illustrated through simulation examples and theoretically predicted graphs (Figures 6, 7 and 8). For reasonably sized datasets, the deviations from a power law size-frequency distribution is not detectable visually. Using a range of values for θ from 0.25 – 0.5, we find that the approximate power law exponent of the combined distribution in 2D differs by 0.15 to 0.3 from that obtained when clustering and linkage are not taken into consideration. Consequently, we find that the difference between the 2D and 1D exponents is only 0.7-0.85 (i.e., < 1.0) for a fault population that contains en echelon clusters and obeys linear displacement-length scaling. If en echelon fault arrays dominate the fault pattern (i.e., $\theta > 0.5$) then the difference $C_2 - C_1$ could be as low as 0.5 according to our simulations. These results show that if clustering and linkage of fault segments is ignored when analysing 1D data, predictions concerning the 2D and full 3D pattern, and/or the number of faults below the resolution of the data, will be meaningless. These results also apply to joint, vein and shear fracture populations even though in our theoretical development we have focused on faults with significant displacements.

Appendix A: Length Distribution of a 1D Sample From a Generalized 2D Population – Assuming Spatial Uniformity

Marrett & Allmendinger (1991) showed the relationship between samples in different dimensions for the specific case of power law size-frequency distributions. Here we derive the analogous results for a general size-frequency distribution. For a given fault length

distribution in 2D, it is of interest to determine the corresponding distribution for faults intersected by a 1D traverse. Let (X, Φ) be a pair of stochastic variables representing respectively the fault length in 2D and the angle between the fault strike and the traverse line. The pdf of (X, Φ) is denoted $f(x, \phi)$, and the event of a fault being intersected by a 1D traverse is denoted I. The pdf of lengths of faults intersected by the traverse differs from the marginal pdf $f(x)$, since small faults are under-sampled in lower dimensions. Let $f(x|i)$ be the pdf of fault lengths for the intersected faults, where we condition on the event I of intersection. The pdf $f(x|i)$ is given by

$$f(x|i) = \frac{f(i,x)}{f(i)} = \frac{1}{f(i)} \int f(i|x,\phi) f(x,\phi)\, d\phi, \tag{15}$$

where $f(i|x, \phi)$ is the probability of intersection conditioned on fault length and orientation and $f(i)$ is the unconditional probability of intersection. Under the assumption of independently, uniformly distributed fault positions, and fault sizes independent of the positions, the probability of a fault being intersected by the traverse is proportional to the fault length perpendicular to the direction of the traverse. This length is given as $X \sin \Phi$. Thus the probability of intersection, conditioned on fault length and orientation, is given as

$$f(i|x,\phi) = kx \sin\phi,$$

where k is a normalization constant. Assuming independence between fault length and orientation the joint pdf can be expressed as $f(x, \phi) = f(x)f(\phi)$. The numerator in Eq. (15) is then given as

$$f(i,x) = \int f(i|x,\phi) f(x) f(\phi)\, d\phi = kx f(x) \int \sin\phi\, f(\phi)\, d\phi = kx f(x) <\sin\phi>, \tag{16}$$

and the unconditional probability of intersection is

$$f(i) = \int f(i,x)\, dx = k <\sin\phi> \int x f(x)\, dx = k <x><\sin\phi> . \tag{17}$$

Substituting (16) and (17) in Eq. (15), we obtain

$$f(x|i) = \frac{kx f(x) <\sin\phi>}{k <x><\sin\phi>} = \frac{x f(x)}{<x>}. \tag{18}$$

Thus the length distribution of intersected faults is independent of the distribution of fault orientation, under the assumption of independence between fault length and orientation. For the power law distribution in 2D, Eq. (1), we have

$$f(x) = \frac{\beta x_0^\beta}{x^{\beta+1}}, \qquad <x> = \frac{\beta x_0}{\beta - 1} \qquad \Rightarrow \qquad f(x|i) = \frac{(\beta-1) x_0^{\beta-1}}{x^\beta},$$

thus the length distribution in 1D is a power law distribution with exponent $\beta - 1$. The minimum fault length x_0 is the same both in 2D and 1D. For the exponential distribution in 2D, Eq. (7), we have

$$f_E(x) = \lambda \exp(-\lambda x), \qquad < x > = \lambda^{-1} \qquad \Rightarrow \qquad f_E(x|i) = \lambda^2 x \exp(-\lambda x),$$

thus the corresponding length distribution in 1D is a gamma distribution $f_G(x) = f_E(x|i)$.

Appendix B: Generalized Derivation of Slope on Bilogarithmic Plots of Fault Number Versus Size

Let $N(x)$ be the number of faults larger than x. This function is typically illustrated by plotting $\log N(x)$ versus $\log x$. The relative cumulative number $N(x)/N_T$, where N_T is the total number of faults, gives an estimate of the cumulative distribution function $S(x)$ defined as

$$S(x) = \int_x^\infty f(x)\, dx$$

where $f(x)$ is the pdf of the length X. For a small number of observed faults, the uncertainty in the estimate $N(x)/N_T$ can be large, as illustrated in Section 1. However, for sufficiently large datasets, the slope of $N(x)$ on log-log plot is approximately equal to the slope of $\log S(x)$ plotted versus $\log x$. This slope is negative, with an absolute value given as

$$C(x) = -\frac{d \log S(x)}{d \log x} = -\frac{d \log S(x)}{dx} \cdot \frac{dx}{d \log x} = -\frac{x}{S(x)} \cdot \frac{dS(x)}{dx} = \frac{x f(x)}{S(x)}.$$

For the power law distribution, Eq. (1), the function $S(x)$ and the slope $C(x)$ are given as

$$S(x) = \left(\frac{x_0}{x}\right)^\beta$$
$$C(x) = x \cdot \frac{\beta x_0^\beta}{x^{\beta+1}} \cdot \left(\frac{x_0}{x}\right)^{-\beta} = \beta,$$

where x_0 is the minimum fault size. Thus the slope $C(x) = \beta$ is constant for all values of x. For the exponential distribution with intensity $\lambda =< x >^{-1}$, Eq. (7), the function $S(x)$ and the absolute value $C(x)$ of the slope are given as

$$S(x) = \exp(-\lambda x)$$
$$C(x) = x \cdot \lambda \exp(-\lambda x) \cdot \exp(\lambda x) = \lambda x,$$

and for the gamma distribution, Eq. (9), the following results are obtained:

$$S(x) = (\lambda x + 1) \exp(-\lambda x)$$
$$C(x) = x \cdot \lambda^2 x \exp(-\lambda x) \cdot \frac{\exp(\lambda x)}{\lambda x + 1} = \frac{(\lambda x)^2}{\lambda x + 1}.$$

Appendix C: Random Linkage Criterion for Faults in 1D

The overlap-to-separation ratio, O/S (Fig. 4), can be used as a linkage criterion for faults in 2D, as discussed in Section 5. However, if there is only 1D data available the separation S is measured, but not the overlap O. Moreover, only the displacements and not the lengths can be measured along a 1D traverse line. Using the relationship $D_{\max} = \gamma L^n$ between length and maximum displacement, and a model for the displacement profile (e.g., Eq. (4)), information from the observed displacements along a traverse can be used to constrain the lengths and thus decide if two faults closely located on the traverse may be classified as linked. The O/S ratio can still not be found exactly, but a statistical approach can be used in the linkage classification. A random linking of faults can be applied, based on the probability of the O/S ratio falling in the specified interval; $\text{Prob}\{M_1 \leq O/S \leq M_2\}$. This probability can then be conditioned on the observed separation; $\text{Prob}\{M_1 \leq O/S \leq M_2 | S = s\}$.

Consider two neighboring faults with the same dip and dip direction along the traverse, with lengths following a power law distribution with exponent b and lower limit of fault length l_0. Assume the intersection points with the traverse are uniformly distributed along the fault traces. The distribution of the overlap O is then found to be

$$f(o) = \begin{cases} \left(\dfrac{b}{b+1}\right)^2 \dfrac{2o}{l_0^2}, & 0 \leq o \leq l_0 \\ \dfrac{2b+1}{(b+1)^2} \cdot \dfrac{2bl_0^{2b}}{o^{2b+1}}, & o \geq l_0 \end{cases} \tag{19}$$

The distribution of the overlap has a power law tail with exponent $2b$ for overlaps larger than the minimum fault length l_0. The conditional probability of $M_1 \leq O/S \leq M_2$ given $S = s$ can be reformulated as $\text{Prob}\{M_1 \leq O/S \leq M_2 | S = s\} = \text{Prob}\{sM_1 \leq O \leq sM_2\}$, which is found to be

$$\text{Prob}\{sM_1 \leq O \leq sM_2\} = \int_{sM_1}^{sM_2} f(o)\, do =$$

$$\begin{cases} \left(\dfrac{b}{b+1}\right)^2 \left(\dfrac{s}{l_0}\right)^2 (M_2^2 - M_1^2), & 0 \leq s \leq \dfrac{l_0}{M_2} \\ 1 - \left(\dfrac{b}{b+1}\right)^2 \left(\dfrac{s}{l_0}\right)^2 M_1^2 - \dfrac{2b+1}{(b+1)^2} \left(\dfrac{s}{l_0}\right)^{-2b} M_2^{-2b}, & \dfrac{l_0}{M_2} \leq s \leq \dfrac{l_0}{M_1} \\ \dfrac{2b+1}{(b+1)^2} \left(\dfrac{s}{l_0}\right)^{-2b} (M_1^{-2b} - M_2^{-2b}), & s \geq \dfrac{l_0}{M_1} \end{cases} \tag{20}$$

Hence, if all parameters M_1, M_2, b and l_0 are known, or estimated, the probability $\text{Prob}\{sM_1 \leq O \leq sM_2\}$ can be found for a given separation $S = s$. The power law expo-

nent of displacements observed along the traverse is $C_1' = \beta_S - 1/n$, see Section 5. The power law exponent of the corresponding lengths is then given by $b = C_1 = nC_1' = n\beta_S - 1$. Thus if n is known (e.g., $n = 1$), the exponent b can be found based on the observations. If the constant γ in the expression $D_{\max} = \gamma L^n$ is known, the lower limit l_0 of fault length can be obtained from a lower limit d_0 of displacements along the traverse.

A random linking criterion can be chosen as "Classify two faults with the same dip and dip direction, separated by a distance s, as linked with probability $\text{Prob}\{sM_1 \leq O \leq sM_2\}$". This can be done by generating a random number u from the uniform distribution on the unit interval $[0, 1]$. If $u \leq \text{Prob}\{sM_1 \leq O \leq sM_2\}$, which occurs with probability $\text{Prob}\{sM_1 \leq O \leq sM_2\}$, the faults are classified as linked. If $u > \text{Prob}\{sM_1 \leq O \leq sM_2\}$, the two observations are treated as isolated faults. Note that this procedure does not give a unique linkage of the faults, only one statistical realization.

The probability $\text{Prob}\{M_1 \leq O/S \leq M_2\}$ can not be related directly to the parameter θ given in Eq. (13). However, based on 2D observations, $\text{Prob}\{M_1 \leq O/S \leq M_2\}$ can empirically be found as

$$\text{Prob}\{M_1 \leq O/S \leq M_2\} = \frac{N_S}{N_I + N_S} \tag{21}$$

where N_I, as before, is the number of isolated faults and N_S is the total number of single fault segments included in linked fault structures. Each linked fault system contains several single fault segments, thus $N_S > N_A$, resulting in $\text{Prob}\{M_1 \leq O/S \leq M_2\} > \theta$.

Acknowledgements

We are grateful to Alberto Malinverno for introducing us to his work on turbidite beds, which Section 3 is based on. We thank Chris Townsend (Statoil R&D) for instigating and supporting this collaborative work, and we thank Henning Omre for reviews and discussions. We also thank the Associate Editor Didier Sornette, and two anonymous reviewers for their constructive comments on our paper. H. Borgos is funded by a PhD grant from VISTA (*Det Norske Videnskaps-Akademi* and *Statoil*). P. Cowie is supported by a Research Fellowship from The Royal Society of London. N. Dawers was funded by NERC-ROPA grant 95/21(GR3/RO43) to J. Underhill and by the Edinburgh Department of Geology and Geophysics.

References

Ackermann, R. V., Schlische, R. W. & Withjack, M. O. (1997). Systematics of an evolving population of normal faults in scaled physical models, *Abstracts with Programs*, number 6 in *29*, Geological Society of America, p. 198.

Aydin, A. & Schultz, R. A. (1990). Effect of mechanical interaction on the development of strike-slip faults with echelon patterns, *Journal of Structural Geology* **12**(1): 123–129.

Berkowitz, B. & Adler, P. M. (1998). Stereological analysis of fracture network structure in geological formations, *Journal of Geophysical Research* **103**(B7): 15339–15360.

Bour, O. & Davy, P. (1999). Clustering and size distributions of fault patterns: Theory and measurements, *Geophysical Research Letters* **26**(13): 2001–2004.

Clark, R. M., Cox, S. J. D. & Laslett, G. M. (1999). Generalization of power-law distributions applicable to sampled fault-trace lengths: model choice, parameter estimation and caveats, *Geophysical Journal International* **136**: 357–372.

Cowie, P. A. (1998). Normal Fault Growth in Three-Dimensions in Continental and Oceanic Crust, *Faulting and Magmatism at Mid-Ocean Ridges*, Geophysical Monograph 106, the American Geophysical Union, pp. 325–348.

Cowie, P. A. & Scholz, C. H. (1992a). Physical explanation for the displacement-length relationship of faults using a post-yield fracture mechanics model, *Journal of Structural Geology* **14**(10): 1133–1148.

Cowie, P. A. & Scholz, C. H. (1992b). Displacement-length scaling relationship for faults: data synthesis and discussion, *Journal of Structural Geology* **14**(10): 1149–1156.

Cowie, P. A., Malinverno, A., Ryan, W. B. F. & Edwards, M. H. (1994). Quantitative fault studies on the East Pacific Rise: A comparison of sonar image techniques, *Journal of Geophysical Research* **99**(B8): 15205–15218.

Davy, P., Sornette, A. & Sornette, D. (1992). Experimental discovery of scaling laws relating fractal dimensions and the length distribution exponent of fault systems, *Geophysical Research Letters* **19**(4): 361–363.

Dawers, N. H. (1996). Fault size distribution on the Volcanic Tableland, California: implications for small-scale fault prediction, *Extended Abstract Volume: Faulting, Fault Sealing and Fluid Flow in Hydrocarbon Reservoirs*, University of Leeds Rock Deformation Research Group, in conjunction with the Petroleum Group & Tectonic Studies Group, The Geological Society, London, pp. 88–89.

Dawers, N. H., Anders, M. H. & Scholz, C. H. (1993). Growth of normal faults: Displacement-length scaling, *Geology* **21**: 1107–1110.

Dawers, N. H. & Anders, M. H. (1995). Displacement-length scaling and fault linkage, *Journal of Structural Geology* **17**(5): 607–614.

Gupta, A. & Scholz, C. H. (2000). A model of normal fault interaction based on observations and theory, *Journal of Structural Geology*. In press.

Hatton, C. G., Main, I. G. & Meredith, P. G. (1993). A comparison of seismic and structural measurements of scaling exponents during tensile subcritical crack-growth, *Journal of Structural Geology* **15**(12): 1485–1495.

Heffer, K. J. & Bevan, T. G. (1990). Scaling Relationships in Natural Fractures: Data, Theory, and Application, *Europec 90*, Society of Petroleum Engineers, pp. 367–376. SPE Paper 20981.

Huggins, P., Watterson, J., Walsh, J. J. & Childs, C. (1995). Relay zone geometry and displacement transfer between normal faults recorded in coal-mine plans, *Journal of Structural Geology* **17**(12): 1741–1755.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2 edn, Wiley.

Malinverno, A. (1997). On the power law size distribution of turbidite beds, *Basin Research* **9**: 263–274.

Marrett, R. (1996). Aggregate properties of fracture populations, *Journal of Structural Geology* **18**(2/3): 169–178.

Marrett, R. & Allmendinger, R. W. (1991). Estimates of strain due to brittle faulting: sampling of fault populations, *Journal of Structural Geology* **13**(6): 735–738.

Marrett, R. & Allmendinger, R. W. (1992). Amount of extension on "small" faults: An example from the Viking graben, *Geology* **20**: 47–50.

Ortega, O. & Marrett, R. (2000). Prediction of macrofracture properties using microfracture information, Mesaverde Group sandstones, San Juan basin, New Mexico, *Journal of Structural Geology* **22**(5): 571–588.

Peacock, D. C. P. & Sanderson, D. J. (1991). Displacements, segment linkage and relay ramps in normal-fault zones, *Journal of Structural Geology* **13**(6): 721–733.

Pickering, G., Bull, J. M. & Sanderson, D. J. (1995). Sampling power-law distributions, *Tectonophysics* **248**: 1–20.

Pickering, G., Bull, J. M. & Sanderson, D. J. (1996). Scaling of fault displacements and implications for the estimation of sub-seismic strain, *in* P. C. Buchanan & D. A. Nieuwland (eds), *Modern Developments in Structural Interpretation, Validation and Modeling*, Geological Society Special Publication No. 99, pp. 11–26.

Piggott, A. R. (1997). Fractal relations for the diameter and trace length of disc-shaped fractures, *Journal of Geophysical Research* **102**(B8): 18121–18125.

Schlische, R. W., Young, S. S., Ackermann, R. V. & Gupta, A. (1996). Geometry and scaling relations of a population of very small rift-related normal faults, *Geology* **24**(8): 683–686.

Scholz, C. H., Dawers, N. H., Yu, J. Z. & Anders, M. H. (1993). Fault growth and fault scaling laws – preliminary results, *Journal of Geophysical Research* **98**(B12): 21951–21961.

Schultz, R. (2000). Fault-population statistics at the Valles Marineris extensional province, Mars: implications for segment linkage, crustal strains, and its geodynamical development, *Tectonophysics* **361**: 169–193.

Segall, P. & Pollard, D. D. (1983). Joint formation in granitic rocks of the Sierra Nevada, *The Geological Society of America Bulletin* **94**: 563–575.

Spryopoulos, C., Griffith, W. J., Scholz, C. H. & Shaw, B. E. (1999). Experimental evidence for different strain regimes of crack populations in a clay model, *Geophysical Research Letters* **26**(8): 1081–1084.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2 edn, Wiley, New York.

Vermilye, J. M. & Scholz, C. H. (1995). Relation between vein length and aperture, *Journal of Structural Geology* **17**(3): 423–434.

Walsh, J. J. & Watterson, J. (1988). Analysis of the relationship between displacements and dimensions of faults, *Journal of Structural Geology* **10**(3): 239–247.

Walsh, J. J., Watterson, J. & Yielding, G. (1991). The importance of small-scale faulting in regional extension, *Nature* **351**: 391–393.

Watterson, J. (1986). Fault dimensions, displacement and growth, *Pure and Applied Geophysics* **124**: 365–373.

Willemse, E. J. M., Pollard, D. D. & Aydin, A. (1996). Three-dimensional analyses of slip distributions on normal fault arrays with consequences for fault scaling, *Journal of Structural Geology* **18**(2/3): 295–309.

Wojtal, S. F. (1996). Changes in fault displacement populations correlated to linkage between faults, *Journal of Structural Geology* **18**(2/3): 265–279.

Part IV

Stochastic model for fault geometry conditioned to seismic data and well observations

Stochastic model for fault geometry conditioned to seismic data and well observations

Hilde G. Borgos

Abstract

A Bayesian framework is used to define a stochastic model of fault patterns conditioned to seismic data and well data. Faults above seismic resolution are considered, enabling assessment of uncertainty for faults detectable from seismic data. The fault pattern is represented through a faulted horizon, with a prior distribution modeling characteristics of fault traces, fault intersections and offset profiles. A likelihood function is defined under the assumption that reflection coefficients of a non-faulted, layered sedimentary reservoir can be modeled as a Gaussian random field with a strong horizontal correlation. Uncertainty in fault patterns is studied through repeated sampling from the posterior distribution.

1 Introduction

Geological faults above seismic resolution can be interpreted from seismic reflection data. Sedimentary rocks tend to form horizontally layered sequences. This layered structure is often observed in the seismic data, where lateral discontinuities visible across succeeding horizontal layers are interpreted as faults. However, the interpretations of faults from seismic data contain uncertainties, both due to the seismic data quality, seismic resolution and the judgment of the interpreter. Furthermore, the subjective human interpretations are not always guaranteed to give geologically realistic fault patterns, especially if the interpreter is poorly trained or has a particular bias. Another problem related to interpreting seismic maps, is the problem of matching faults interpreted from succeeding two dimensional maps to create a three dimensional fault pattern. Badley et al. (1990) discuss a methodology for interpreting faults from three dimensional seismic data.

Attempts have been made to create algorithms for detecting faults from three dimensional (3D) seismic data. Bahorich and Farmer (1995) introduce a successful technique (the Coherence Cube) for tracking/highlighting faults from 3D seismic data. The method is essentially a moving window correlation algorithm, and is patented, see Bahorich and

Farmer (1996). Further developments based on the idea behind the Coherence Cube are presented by Marfurt et al. (1998) and Marfurt et al. (1999). Two alternative techniques for detecting faults and stratigraphic boundaries from 3D seismic data are described by Luo et al. (1996). The techniques presented in the cited papers are deterministic procedures used at an early stage in the interpretation process, and thus making the interpretation easier and faster. The techniques give indications of fault positions, lengths, connectivity and orientations. Experience and use of these techniques have shown that they neither improve fault resolution, nor do they significantly reduce uncertainty. They also tend to be prone to seismic noise and only have a real benefit if used at an early stage.

In this work, the problem of detecting faults from 3D seismic data is considered from a statistical point of view. The method is both novel and clearly different from the previously mentioned correlation techniques. A stochastic model for post-sedimentary fault patterns is defined, conditioned to 3D seismic data and well observations. The layered structure observed in seismic data is incorporated in the model, and the aim is to locate faults based on discontinuities in the lateral consistency. Additional information about the reservoir obtained from wells are also included in the model. Furthermore, general geological knowledge of faults and fault patterns are incorporated in the model in an attempt to produce geologically realistic fault patterns.

Previous work on stochastic modeling of faults and fractures include both pixel or grid based approaches, see for example Gringarten (1996, 1998), object based models, e.g., Chilès (1988), Brand and Haldorsen (1988), Gauthier and Lake (1993), Munthe et al. (1993), Munthe et al. (1994), Wen and Sinding-Larsen (1997), and one dimensional models where faults or fractures only appear as points, see Belfield (1998). Some of the cited papers include information from observations. Fault or fracture observations in wells are conditioned to in some cases, while often data is only used to estimate parameters related to characteristics of faults and fault patterns. For example, seismic data or outcrop data are used to estimate distributions of fault size or orientation, and spatial fault densities. Previous stochastic modeling of faults based on seismic observations mainly use interpreted seismic data to model sub-seismic fault patterns. In contrast, the stochastic model described here is defined for faults at seismic scales, and is used to assess uncertainty in the fault pattern at these scales.

A Bayesian framework is used to present the stochastic model for fault patterns, and to condition to available information like seismic data and well observations. For previous work using a Bayesian framework in reservoir modeling, see for example Lia et al. (1997), Omre and Tjelmeland (1997), Eide et al. (1997a), Eide et al. (1997b), Eide (1999). In this work the focus is on the network of faults in the reservoir. The relationship between specific reservoir characteristics and the available observations is integrated in the model, enabling conditioning of fault patterns to the observations. General geological knowledge is incorporated in a prior distribution, while the available observations are integrated in the model through the likelihood function. The posterior distribution combines the general

knowledge and the observations. Assessment of uncertainty of fault geometries can be based on repeated sampling from the posterior model.

Both the prior model and the likelihood function presented in this work are defined using random fields: Markov Random Fields and Gaussian Random Fields, see Cressie (1993), Tjelmeland (1996), Abrahamsen (1997).

2 Bayesian Framework

A stochastic model is defined for a faulted horizon H, a topographic surface embedded in three dimensions. Faults, their positions, orientations and displacements, are visible as discontinuities in the surface. Let $\boldsymbol{O}$ denote available observations obtained from the reservoir under study, consisting of seismic data and well observations. The aim of this work is to generate geological fault patterns, represented by faulted horizons H, conditioned on the available observations $\boldsymbol{O}$. A Bayesian framework is used to define the relationship between the faulted horizon and the observations. A prior model $f(h)$ of the faulted horizon includes general geological knowledge about faults and fault patterns, and is discussed in Section 3. The relationship between the faulted horizon and the observations is defined through the likelihood function $f(\boldsymbol{o}|h)$, discussed in Section 4. Combining the prior distribution and the likelihood function, the posterior distribution is obtained:

$$f(h|\boldsymbol{o}) = \text{const} \times f(h)f(\boldsymbol{o}|h). \tag{1}$$

The posterior distribution combines general geological knowledge and reservoir specific observations and is discussed in Section 5. Sampling from the posterior distribution is described in Section 6.

Some basic model assumptions are made. The vertical extents of the faults are assumed to be large compared to the vertical extent of the region under study, thus vertically all faults penetrate completely through the region. Furthermore a crude approximation is made, assuming that faults within the region has a vertical dip and a constant offset vertically.

3 Prior Model of Geological Fault Patterns

Post sedimentary faults are studied in this work. Consider a layered rock where all layers originally are horizontal, as is often the case for sedimentary rocks. This corresponds to the fundamental geological principle that sedimentary rocks are deposited horizontally, see Monroe and Wicander (1994, chap. 17). When post sedimentary faults emerge through the rock volume, rock in different parts of a layer are displaced with different offsets, and

an originally horizontal surface is deformed into a topographic surface embedded in three dimensions, denoted a horizon.

Faults form brittle, discrete deformation in sedimentary strata. Other types of deformation, like ductile deformation, folding and tilting, are in this work assumed to be relatively small and are modeled as natural variations in the topographic surface. Thus the variations in the horizon are composed of displacements caused by faulting and other natural variations.

When numerous faults are present in an area, a fault pattern is created. A fault pattern is usually best observed by examining a single horizon from above, i.e., a topographic map. However, the total three dimensional extent of the faults can not be obtained from the surface, and only the pattern of fault traces on the horizon can be mapped. Fault traces are the intersections of fault planes with the horizon. The fault pattern can be represented through the number of faults, their position, geometry, displacement and displacement direction.

Under the model assumptions outlined in Section 2, a series of overlying horizons will have similar topographies, and the fault pattern can be represented through a single representative horizon H. The prior model of the geological fault pattern is defined through the probability density function (pdf) $f(h)$ of the faulted horizon H, and should contain general geological knowledge about faults and fault patterns. In Section 3.1 the faulted horizon is parameterized by an image, and in Section 3.2 a prior model of the image is defined.

3.1 Faulted horizon

A fault pattern is represented by the topography of a faulted horizon H, where faults are recognized as discontinuities in the surface. The number of faults, their position and the offsets are found from these discontinuities. The faulted horizon is parameterized by an image, where both pixels, edges and vertices are included in the model. Figure 1 gives an illustration of two types of pixels and the corresponding edges and vertices.

Figure 1: (a) Square pixel and (b) hexagonal pixel. The pixels are colored gray, the edges around the pixel are marked with solid lines and the vertices with circles.

Edges form an image of fault traces on a horizon. Each edge E is assigned a value in the set $\Omega_e = \{0, 1\}$, and fault traces are composed of connected edges of values $E = 1$.

Vertices represent the offsets along fault traces. Vertices V along connected edges of values $E = 1$ are assigned a value corresponding to the offset of the fault at the point of the vertex. A vertex takes a value in the set $\Omega_v \subseteq \mathbb{R}$, where the sign of the offset indicates the offset direction. Vertices connecting only edges of value $E = 0$ is assigned zero offset $V = 0$.

Pixels represent the topography of the faulted horizon, consisting of displacements caused by faulting and other natural variations. The value of a pixel P determines the altitude of the surface in the center point of the pixel. Each pixel is assigned a value from the set $\Omega_p \subseteq \mathbb{R}$.

Let $\mathcal{E}$, $\mathcal{V}$ and $\mathcal{P}$ denote the index sets of edges, vertices and pixels respectively, and let $n_e = |\mathcal{E}|$, $n_v = |\mathcal{V}|$ and $n_p = |\mathcal{P}|$ be the number of edges, vertices and pixels. The edge values are given by the vector $\boldsymbol{E} = \{E_i;\, i \in \mathcal{E}\}$, the vertices by $\boldsymbol{V} = \{V_i;\, i \in \mathcal{V}\}$ and the pixels by $\boldsymbol{P} = \{P_i;\, i \in \mathcal{P}\}$. All edges E_i, vertices V_i and pixels P_i are treated as stochastic variables, and the sample space of the image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$ is given by $\Omega = \Omega_e^{n_e} \times \Omega_v^{n_v} \times \Omega_p^{n_p}$.

A hexagonal grid is used to represent the image, as was used by Besag (1989) and Tjelmeland and Besag (1998). An example of a hexagonal grid is given in Figure 2. The main advantage of a hexagonal grid, compared to a square grid, is that it gives a greater flexibility in drawing angles of different sizes between fault traces.

The number of pixels, edges and vertices are easily found in the hexagonal grid. Let r be the number of pixels along one side of the hexagonal grid. This is equal to the number of concentric hexagonal bands of pixels included in the grid, starting with the center pixel as band number one. The total number of pixels, edges and vertices in the grid is $n_p = 3r(r-1)+1$, $n_e = 3r(3r-1)$ and $n_v = 6r^2$ respectively.

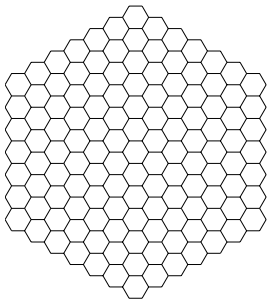


Figure 2: The image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$ is represented by a hexagonal grid.

The image composed of the variables $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$ gives a discrete parameterization of the topography of the faulted horizon. A continuous representation of the horizon is obtained by triangulation between pixels, where faults are represented as discontinuities in the triangulation surface. The triangulation is uniquely determined by the stochastic variables $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$. The resulting topographic surface is also stochastic, and is denoted $H = H(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$. The altitude of the surface H at a point (x, y) is given by $H(x, y)$. At the center point (x, y) of any pixel, with value $P = p$, the altitude of the surface equals the value of the pixel, $H(x, y) = p$.

The first step in the triangulation is to draw triangles between the center points of the pixels, as shown in Figure 3. Each side of a triangle crosses an edge in the image. If one or more of the edges have value $E = 1$, a fault intersects the triangle. The triangle is then divided into smaller triangles in order to obtain a discontinuity along the fault trace, as shown in Figure 4. In this way the image is covered by triangles, where the surface $H(x, y)$ is continuous within each triangle, and discontinuities in the surface are all located along triangle sides. An example of a topographic surface, found by triangulation from an image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$, is shown in perspective in Figure 5.

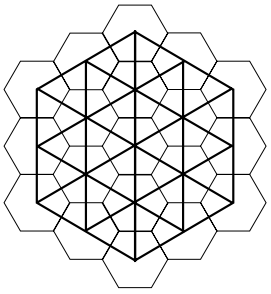


Figure 3: As a first step in the triangulation, triangles are drawn between the center points of the pixels.

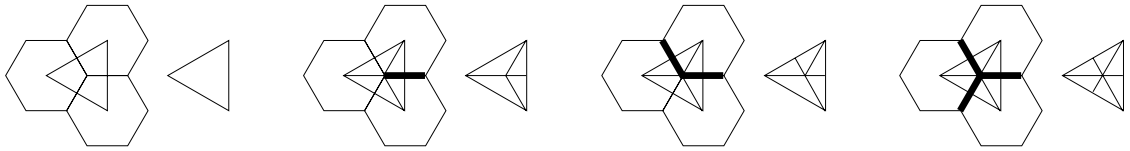


Figure 4: Subdivisions of the triangles shown in Figure 3. The pixels are shown to the left in each figure, and edges of value $E = 1$ are marked with thick solid lines. The subdivisions of the triangle are illustrated to the right. (a) If no faults intersect the triangle, no subdivision is made. (b) The end of a fault trace intersects the triangle, and the triangle is divided into 4 smaller triangles. (c) If a fault trace crosses the triangle, 5 sub-triangles are constructed. (d) The triangle contains an intersection of fault traces, and is divided into 6 smaller triangles.

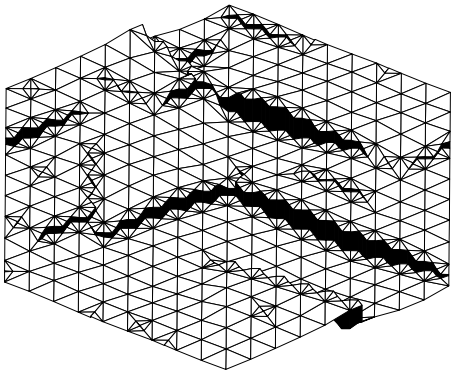


Figure 5: Perspective plot of the topography of the horizon H, constructed by triangulation of the pixels in an image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$. The fault planes, which are all vertical, are colored dark.

3.2 Prior model

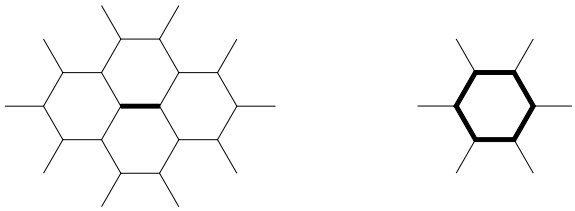
The prior model of a faulted horizon should contain general geological knowledge about faults and fault patterns. The prior model of the edges $\boldsymbol{E}$ should thus contain information on typical fault patterns in a horizon, the prior model of the vertices $\boldsymbol{V}$ should reflect characteristics of fault offsets, and the prior model of the pixels $\boldsymbol{P}$ should produce a satisfactory topography of the faulted horizon. Combining the prior distributions of the edges, vertices and pixels, a prior model of the faulted horizon H is obtained. The pdf $f(h)$ of H can be expressed as follows:

$$f(h) = f(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{p}) = f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v}) f(\boldsymbol{v}|\boldsymbol{e}) f(\boldsymbol{e}). \tag{2}$$

3.2.1 Fault traces

The edges $\boldsymbol{E}$ form an image of fault traces on a horizon, and the prior model of fault traces is based on the work by Tjelmeland and Besag (1998). They consider Markov random fields containing higher order interactions, where a Markov random field is defined on the pixels in the image. In this work a corresponding model is used for the edges. The aim is to include higher order interactions between edges in the image, and to control, to some extent, the pattern of fault traces. The image is represented by a hexagonal grid, as described in Section 3.1. Typical angles of intersection between two fault traces are angles of approximately $60°$, which are easily modeled in the hexagonal grid, but angles of around $90°$ are also common and can be modeled using this grid. Edge effects are accounted for by using circular boundary conditions.

A neighborhood system must be chosen to define the Markov random field used to model



(a) (b)

Figure 6: (a) The neighborhood of the edge in the center, drawn with a thick line, is composed of the edges drawn with thin lines. (b) The maximum clique of the neighborhood in (a). The edges drawn with thick lines form the inner edges, and the edges drawn with thin lines form the outer edges.

the fault traces. The neighborhood illustrated in Figure 6a is used, and the corresponding maximum clique is shown in Figure 6b. The maximum clique consists of the largest set of edges where all pairs of edges are neighbors. The edges contained in the maximum clique are classified as inner and outer edges, where inner edges are drawn with thick lines in Figure 6b, and outer edges with thin lines.

According to the Hammersley-Clifford theorem, see for example Winkler (1995), the distribution of any Markov random field $\boldsymbol{E}$ can be expressed as a Gibbs distribution, see Geman and Geman (1984). The pdf of the Gibbs distribution is given as

$$f(\boldsymbol{e}) = \text{const} \times \exp\{-\sum_{c \in \mathcal{C}} \omega_c(\boldsymbol{e}_c)\}, \tag{3}$$

where $\mathcal{C}$ is the set of all cliques and $\omega_c(\boldsymbol{e}_c)$ is the potential function of the edges $\boldsymbol{e}_c$ contained in the clique c. The sum $\sum_{c \in \mathcal{C}} \omega_c(\boldsymbol{e}_c)$ is called the energy function. Using the neighborhood shown in Figure 6a, possible cliques are the maximum clique shown in Figure 6b and all subsets of the maximum clique. Only the maximum cliques are assigned non-zero potentials in this work. Thus the pdf of the edges $\boldsymbol{E}$ can be expressed as

$$f(\boldsymbol{e}) = \text{const} \times \exp\{-\sum_{c \in \mathcal{C}^m} \omega_c(\boldsymbol{e}_c)\}, \tag{4}$$

where $\mathcal{C}^m \subset \mathcal{C}$ is the set of all maximum cliques. The maximum clique $\boldsymbol{e}_c$ contains 12 edges, where each edge takes a value in the set $\Omega_e = \{0, 1\}$. Thus there are $2^{12} = 4096$ different clique configurations, with a corresponding number of function values $\omega_c(\boldsymbol{e}_c)$. In order to reduce the number of potentials that needs to be specified, the clique configurations are classified into a limited number of classes and all members of a class are assigned equal potentials. A possible classification is shown in Figure 7. Rotations and reflections of the configurations belong to the same class, and are omitted in the figure. The clique configurations in the figure are classified as background, point, end, short-line, line, two-lines, angle, single-cross, double-cross or edge-background. The latter is the class of configurations where all inner edges have value 0, but at least one of the outer edges has a value

of 1. All configurations not shown in Figure 7, or not being a rotation or reflection of a configuration in the figure, are classified as other.

The pdf in Expression (4) is defined by specifying potentials $\omega_c(\boldsymbol{e}_c) = \gamma_i$ for the 11 classes of clique configurations. The potentials are given in Table 1. Since any constant can be added or subtracted in the exponent in Expression (4) without changing the prior pdf $f(\boldsymbol{e})$, the critical choice is not the actual potential values, but the differences between the potentials. Thus one class can be assigned a zero potential, and all other potentials are defined relative to this class. The class of edge-background is chosen as a zero potential reference class.

The classification of cliques shown in Figure 7, with the potentials given in Table 1, defines an isotropic distribution of the orientation of fault traces. The configuration classes can be subdivided to produce anisotropic distributions of fault orientation, assigning different potentials to fault traces depending on the direction of the trace. An example of a subdividing of the configuration class line is shown in Figure 8. Directions 1, 3 and 5 are rotations and reflections of the configuration in Figure 7g, while directions 2, 4 and 6 are reflection and rotations of the configuration in Figure 7h.

Subdividing of the configuration classes end, short-line and two-lines can also be applied, based on reflections and rotations of the configurations in Figures 7c-d, 7e-f and 7i-j respectively. The configuration classes angle, single-cross and double-cross contain two or more fault traces of different orientations, and subdividing according to directions is not applied to these classes. Furthermore, neither the configuration classes point, other nor edge-background are subdivided. The total number of potentials are listed in Table 2. The potentials of the subdivided classes are donated $\gamma_{i,j}$, $j = 1, \ldots, 6$, where j gives the direction of the fault trace.

Configuration name	*Configuration* $\boldsymbol{e}_c$	$\omega_c(\boldsymbol{e}_c)$
Background	Fig. 7a	γ_1
Point	Fig. 7b	γ_2
End	Fig. 7c, d	γ_3
Short-line	Fig. 7e, f	γ_4
Line	Fig. 7g, h	γ_5
Two-lines	Fig. 7i, j	γ_6
Angle	Fig. 7k	γ_7
Single-cross	Fig. 7l	γ_8
Double-cross	Fig. 7m	γ_9
Other	Not illustrated	γ_{10}
Edge-background	Fig. 7n	0.0

Table 1: Potentials $\omega_c(\boldsymbol{e}_c) = \gamma_i$ for the configuration classes shown in Figure 7.

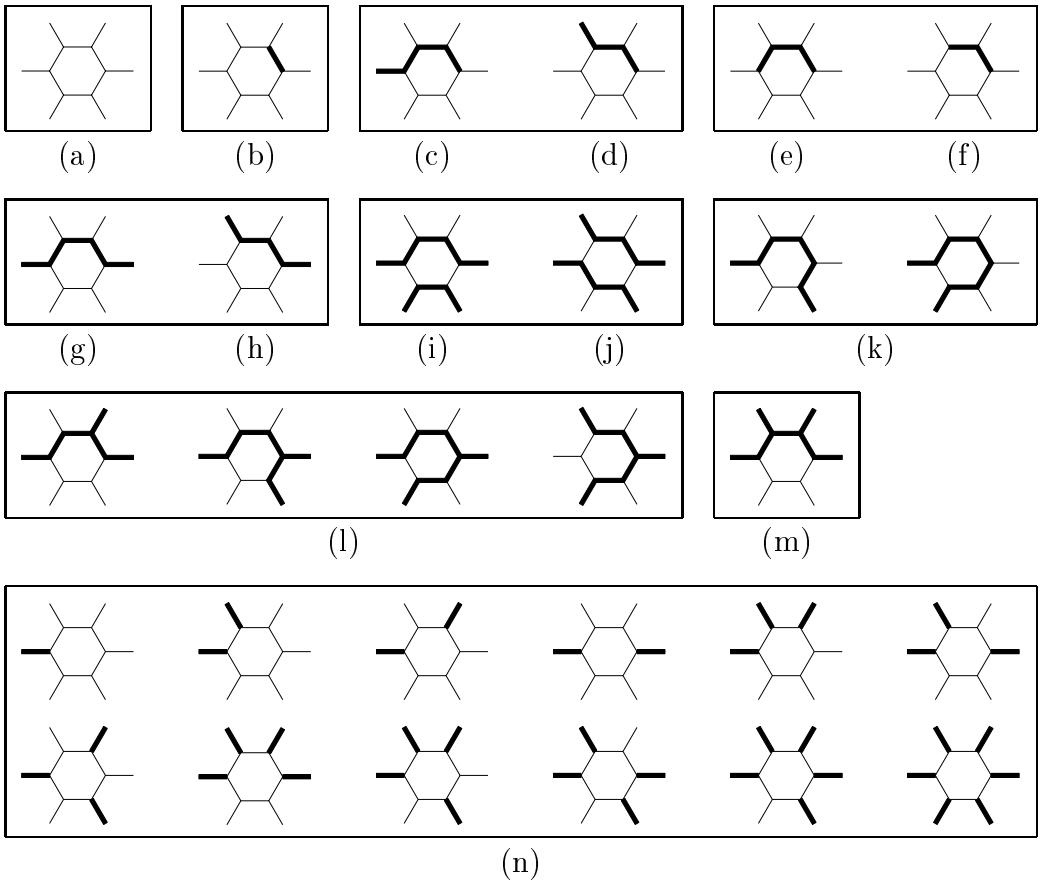


Figure 7: Classes of clique configurations, up to rotation and reflection. Thick solid edges have value $E = 1$, others have value $E = 0$. The clique configurations are classified as (a) background, (b) point, (c)–(d) end, (e)–(f) short-line, (g)–(h) line, (i)–(j) two-lines, (k) angle, (l) single-cross, (m) double-cross and (n) edge-background.

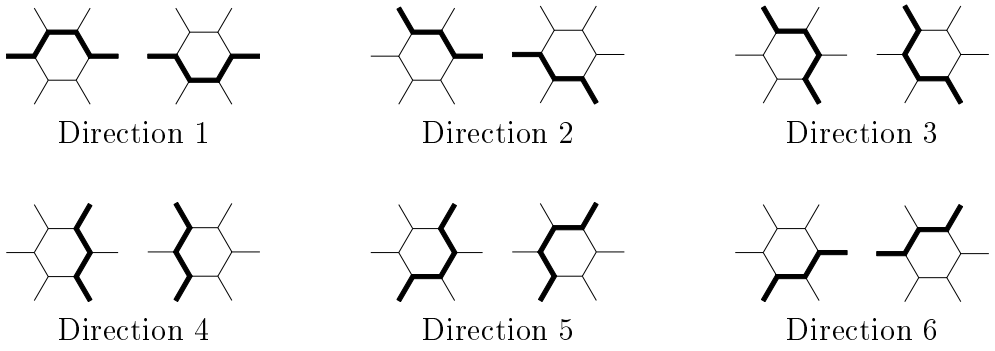


Figure 8: Subdividing of the configuration class line into 6 classes of different orientations.

Configuration name	*Configuration* $\boldsymbol{e}_c$	$\omega_c(\boldsymbol{e}_c)$
Background	Fig. 7a	γ_1
Point	Fig. 7b	γ_2
End	Fig. 7c	$\gamma_{3,1}$, $\gamma_{3,3}$, $\gamma_{3,5}$
	Fig. 7d	$\gamma_{3,2}$, $\gamma_{3,4}$, $\gamma_{3,6}$
Short-line	Fig. 7e	$\gamma_{4,1}$, $\gamma_{4,3}$, $\gamma_{4,5}$
	Fig. 7f	$\gamma_{4,2}$, $\gamma_{4,4}$, $\gamma_{4,6}$
Line	Fig. 7g	$\gamma_{5,1}$, $\gamma_{5,3}$, $\gamma_{5,5}$
	Fig. 7h	$\gamma_{5,2}$, $\gamma_{5,4}$, $\gamma_{5,6}$
Two-lines	Fig. 7i	$\gamma_{6,1}$, $\gamma_{6,3}$, $\gamma_{6,5}$
	Fig. 7j	$\gamma_{6,2}$, $\gamma_{6,4}$, $\gamma_{6,6}$
Angle	Fig. 7k	γ_7
Single-cross	Fig. 7l	γ_8
Double-cross	Fig. 7m	γ_9
Other	Not illustrated	γ_{10}
Edge-background	Fig. 7n	0.0

Table 2: Potentials $\omega_c(\boldsymbol{e}_c)$ when some classes are subdivided to define potentials depending on the orientation of the fault traces. The subscript j of $\gamma_{i,j}$ indicates the direction of the fault trace.

Figures 9 and 10 show independently generated realizations from $f(\boldsymbol{e})$, using the different potentials listed in Table 3. Simulation from the distribution $f(\boldsymbol{e})$ is described in Appendix A, Algorithm 2. Several simulations are run using different potential values, to study the variations in the prior distribution. The number of concentric bands of pixels are 9, resulting in 217 pixels, 702 edges and 486 vertices. Examples 1–4, Figure 9, use isotropic prior distributions of $\boldsymbol{E}$. In examples 5–8, Figure 10, fault traces with orientation in direction NW-SE are favored, which is direction 2 in Figure 8.

In example 1 angles and single-crosses are assigned low potentials, creating fault patterns with a number of bending and intersecting faults. In examples 2–4 the potentials of angles and single-crosses are increased, resulting in fault patterns of isolated, straight fault traces. The pdf used in example 3 has a lower potential for points and short-lines than in example 2, and the fault traces tend to be shorter. In example 4 the potential of points and short-lines is decreased even further. The result is a fault pattern containing a relatively large number of short fault traces.

The realizations in examples 5–8 are generated from anisotropic prior distributions. In examples 6 and 7 the fault patterns are dominated by a number of long fault traces in direction NW-SE. This characteristic is hardly detectable in example 5 where the potentials of direction NW-SE are slightly lower than potentials of the other directions, but angles and single-crosses are assigned low potentials. Direction NW-SE dominates in the realizations of example 8, as it does in examples 6 and 7. However, due to a higher potential for lines,

(a)

Example No.	Potentials		
	γ_1	γ_9	γ_{10}
1–8	-0.1	1.1	3.4

(b)

Example No.	Potentials						
	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
1	1.6	-0.1	1.4	-0.2	1.025	0.0	0.1
2	1.8	-0.1	1.6	-0.2	1.025	0.2	0.9
3	1.6	-0.1	1.4	-0.2	1.025	0.5	0.9
4	1.2	-0.1	1.0	-0.2	1.025	0.5	0.9

(c)

Example No.	Potentials										
	γ_2	$\gamma_{3,2}$	$\gamma_{3,j}$	$\gamma_{4,2}$	$\gamma_{4,j}$	$\gamma_{5,2}$	$\gamma_{5,j}$	$\gamma_{6,2}$	$\gamma_{6,j}$	γ_7	γ_8
5	1.6	-0.2	-0.1	1.3	1.4	-0.25	-0.2	0.925	1.025	0.0	0.1
6	1.6	-0.2	-0.1	1.3	1.4	-0.275	-0.2	0.925	1.025	0.0	0.1
7	1.6	-0.2	-0.1	1.3	1.4	-0.3	-0.2	0.925	1.025	0.0	0.1
8	1.6	-0.2	-0.1	1.3	1.4	-0.2	-0.1	0.925	1.025	0.5	0.9

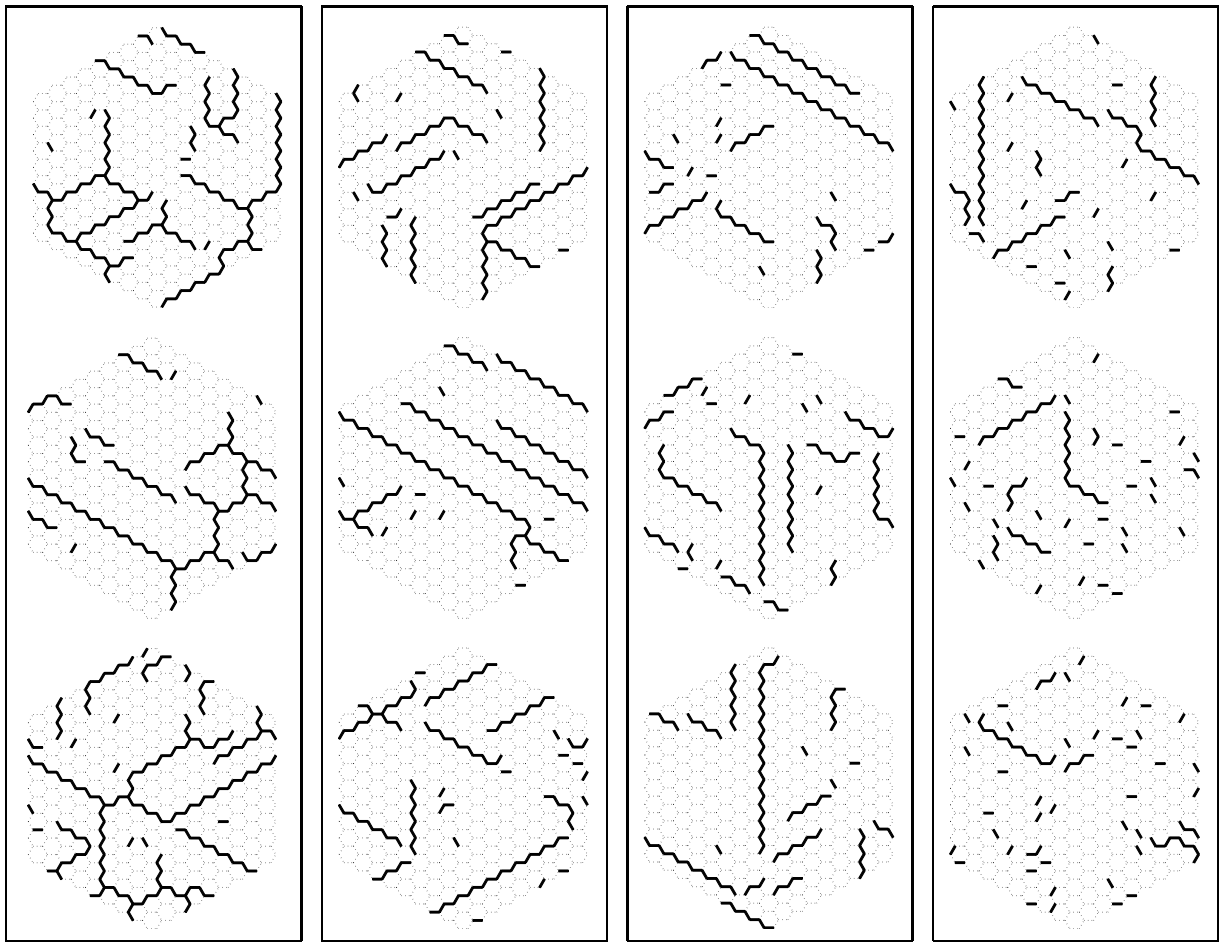
Table 3: Potentials used in the prior pdf $f(\boldsymbol{e})$. The potentials in (a) are kept constant in all examples. The potentials in (b) give an isotropic prior distribution while the potentials in (c) give an anisotropic prior pdf which favors fault orientations in direction 2, see Figure 8. In the potentials $\gamma_{i,j}$, $i = 3, \ldots, 6$, j takes the values $j = 1, 3, \ldots, 6$.

angles and single-crosses in example 8, the faults are shorter and few intersections are observed.

3.2.2 Fault offsets

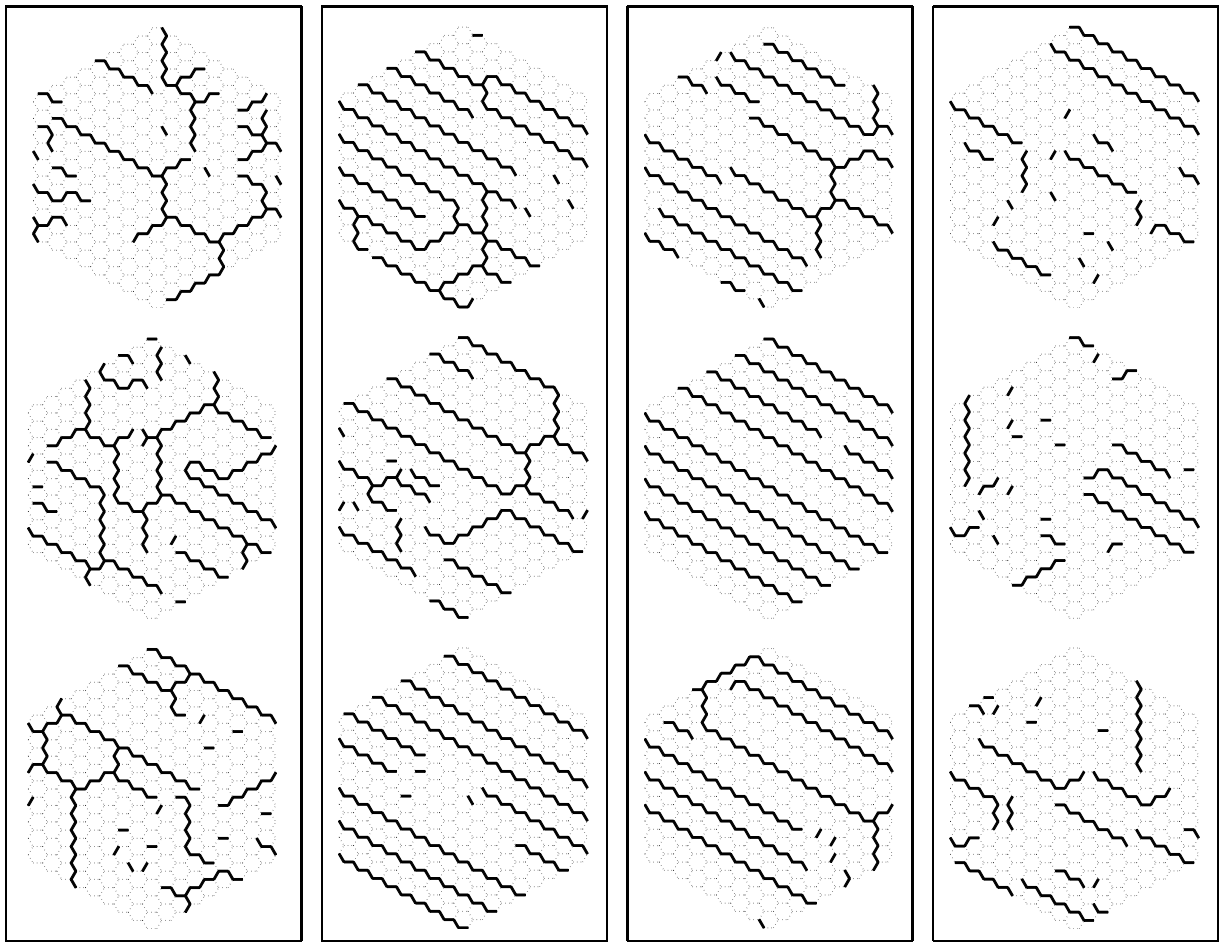
The vertices $\boldsymbol{V}$ represent the offsets along fault traces, where the offset tends to vary along a trace. The maximum offset depends on the horizontal extent, or length, of the fault trace, and is typically located near the center point of the trace. Faults with long trace lengths tend to have a larger maximum offset than faults with short trace lengths. A number of geological studies suggest a relationship $D \propto L^{\alpha}$ between maximum displacement D and length L, see for example Walsh and Watterson (1987, 1988), Gauthier and Lake (1993), Dawers et al. (1993).

Fault traces are represented as connected edges of value $E = 1$. A fault trace can have an isolated fault tip, where one end of an edge of value $E = 1$ is only connected to edges of value $E = 0$, as in the clique configurations in Figure 7c-d. Alternatively, the fault trace can terminate in another fault trace, producing a single-cross as in the configurations in Figure 7l, or a double-cross as shown in Figure 7m. The number of faults and their traces



Example 1 Example 2 Example 3 Example 4

Figure 9: Realizations of $\boldsymbol{E}$ from $f(\boldsymbol{e})$, using different potentials with isotropic distribution of orientation. The potentials are given in Table 3.



Example 5 Example 6 Example 7 Example 8

Figure 10: Realizations of $\boldsymbol{E}$ from $f(\boldsymbol{e})$, using different potentials with anisotropic distribution of orientation. Direction NW-SE is favored. The potentials are given in Table 3.

are determined by the realization of the edges. The circular boundary conditions used in the prior model of $\boldsymbol{E}$ are adopted.

Let $b \subset \mathcal{V}$ be an index set consisting of all vertices along a fault trace, and let $\mathcal{B}_e = \{b_1, b_2, \ldots, b_n\}$ be the set of all fault traces produced by the edge values $\boldsymbol{E} = \boldsymbol{e}$. Each vertex is defined as belonging to a maximum of one fault trace, thus $b_j \cap b_k = \emptyset$ for any $b_j, b_k \in \mathcal{B}_e$. At intersection points of two traces it is not uniquely defined which two edges belong to the same trace, while the third edge is part of the intersecting trace. Thus for each realization $\boldsymbol{e}$ of edges there is a number, $n_b(\boldsymbol{e})$, of possible ways of defining the set $\mathcal{B}_e$. All these partitionings are assigned equal probability $n_b(\boldsymbol{e})^{-1}$. The index set of all vertices not included in any fault trace is denoted $b_0 = \mathcal{V} \setminus \{\cup_{b_j \in \mathcal{B}_e} b_j\}$, where for any vertex V_i, $i \in b_0$, the common value of all edges connected to the vertex is $E = 0$.

Consider a fault trace $b \in \mathcal{B}_e$ composed of n connected edges of value $E = 1$. The fault trace is classified according to the number of isolated fault tips, $i \in \{0, 1, 2\}$, and is assigned a standardized offset profile $g_i(x, n)$, $x \in [0, 1]$. The offset profile has a maximum at the center point of the fault trace, descending to zero at isolated fault tips. The offset profiles are shown in Figure 11 and are defined as

$$\begin{aligned} g_0(x,n) &= \mu(n) \\ g_1(x,n) &= \begin{cases} 4\mu(n)x(1-x), & x \leq 0.5 \\ \mu(n), & x > 0.5 \end{cases} \\ g_2(x,n) &= 4\mu(n)x(1-x), \end{aligned} \tag{5}$$

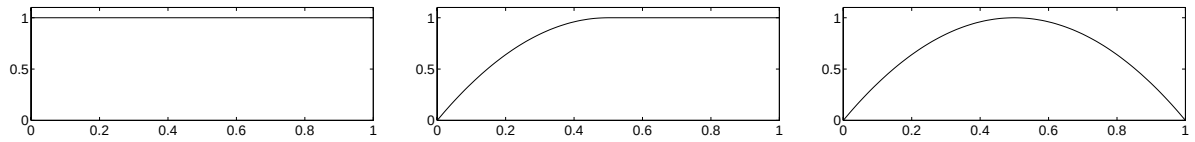
where the function $\mu(n)$ is used to model the relationship between the maximum offset and the length n of the fault trace, and is chosen as $\mu(n) = \mu_v n$ where μ_v is a constant. A discrete representation of the offset profile $g_i(x, n)$ is given by the vector $\boldsymbol{g}_{i,n}$, with one component for each vertex along b. The offset values of vertices along the trace are given as $\boldsymbol{V}_b = D_b \boldsymbol{g}_{i,n}$, where the stochastic variable D_b is the offset direction of fault b, with two possible values $D_b \in \{-1, 1\}$. The pdf of the vertex values $\boldsymbol{V}_b$ conditioned on the offset direction D_b is given by $f(\boldsymbol{v}_b | d_b) = \delta(\boldsymbol{v}_b - d_b \boldsymbol{g}_{i,n})$, where $\delta(t)$ is the Dirac delta function. The prior distribution of D_b is chosen as $f(1) = \theta$, $f(-1) = 1 - \theta$, where $0 \leq \theta \leq 1$, resulting in the following pdf of $\boldsymbol{V}_b$:

$$f(\boldsymbol{v}_b) = \sum_{d_b \in \{-1,1\}} f(\boldsymbol{v}_b | d_b) f(d_b) \; = \; \theta \delta(\boldsymbol{v}_b - \boldsymbol{g}_{i,n}) + (1 - \theta)\delta(\boldsymbol{v}_b + \boldsymbol{g}_{i,n}). \tag{6}$$

When the offset direction D_b is known, the vertex values $\boldsymbol{V}_b$ are deterministically given. Thus the distribution of the offset along the fault trace b is uniquely defined through the dip parameter θ, the length n of the trace and the size parameter μ_v.

The vertices in b_0 represent points not included in any fault traces. These points have zero offset, implying vertex values of $V_i = 0$, $i \in b_0$. The vertices in b_0 are denoted $\boldsymbol{V}_{b_0} = \{V_i; i \in b_0\}$, and have pdf

$$f(\boldsymbol{v}_{b_0}) = \delta(\boldsymbol{v}_{b_0}). \tag{7}$$



$g_0(x,n)$ $g_1(x,n)$ $g_2(x,n)$

Figure 11: Functions $g_i(x,n)$ representing the standardized offset profiles along a fault trace of i isolated fault tips.

The pdf of all vertices $\boldsymbol{V}$, conditioned on the edges $\boldsymbol{E}$, is obtained by combining the pdfs (6) and (7):

$$f(\boldsymbol{v}|\boldsymbol{e}) = n_b(\boldsymbol{e})^{-1} f(\boldsymbol{v}_{b_0}) \prod_{b \in \mathcal{B}_e} f(\boldsymbol{v}_b), \tag{8}$$

where vertices $\boldsymbol{V}_b$ and $\boldsymbol{V}_{b'}$ are assumed to be independent for any two fault traces $b, b' \in \mathcal{B}_e$. The number of terms in the product depends on the dimension $|\mathcal{B}_e|$ of the set $\mathcal{B}_e$, which is determined by the edge values $\boldsymbol{e}$. The non-zero values of the pdf can be expressed as

$$f(\boldsymbol{v}|\boldsymbol{e}) = n_b(\boldsymbol{e})^{-1} \theta^k (1-\theta)^{|\mathcal{B}_e|-k}, \qquad 0 \le k \le |\mathcal{B}_e|, \tag{9}$$

where k is the number of faults $b \in \mathcal{B}_e$ offset in direction $D_b = 1$.

3.2.3 Topography of faulted surface

The pixels $\boldsymbol{P}$ represent the topography of the faulted horizon. The pattern of fault traces on the horizon and the offsets along the traces are represented by the edges $\boldsymbol{E}$ and vertices $\boldsymbol{V}$ respectively. The pixel values $\boldsymbol{P}$ should produce an image of a surface, reproducing the discontinuities and offsets along the fault traces. Note that absolute depth is not defined, but the pixel values give relative altitudes of the topography of the faulted surface. Edge effects are accounted for by using free boundary conditions.

A multivariate Gaussian distribution is used as a prior model for pixel values. The mean value μ should depend on the edge values $\boldsymbol{E} = \boldsymbol{e}$ and vertex values $\boldsymbol{V} = \boldsymbol{v}$, thus $\mu = \mu(\boldsymbol{e}, \boldsymbol{v})$, while the covariance matrix Σ_p is defined independently of $\boldsymbol{e}$ and $\boldsymbol{v}$. The prior pdf of $\boldsymbol{P}$ is then given as the pdf $\phi(\cdot)$ of the multivariate Gaussian distribution:

$$\begin{aligned} f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v}) &= \phi(\boldsymbol{p}; \mu(\boldsymbol{e},\boldsymbol{v}), \Sigma_p) \\ &= (2\pi)^{-n_p/2} |\Sigma_p|^{-n_p/2} \exp\{-(\boldsymbol{p}-\mu(\boldsymbol{e},\boldsymbol{v}))' \Sigma_p^{-1} (\boldsymbol{p}-\mu(\boldsymbol{e},\boldsymbol{v}))/2\}. \end{aligned} \tag{10}$$

The mean values $\mu(\boldsymbol{e},\boldsymbol{v})$ should reflect the discontinuities along fault traces, with the correct offset values across the discontinuities. Consider two adjacent pixels P_i and P_j, where E_{ij} is the edge between the pixels and V_{ij1} and V_{ij2} the vertices at the ends of the

edge, see Figure 12. Let $\bar{V}_{ij}$ be the average of the two vertex values. For a given realization $(\boldsymbol{E}, \boldsymbol{V}) = (\boldsymbol{e}, \boldsymbol{v})$, denote by $\mu(\boldsymbol{e}, \boldsymbol{v}) = (\mu_1, \mu_2, \ldots, \mu_{n_p})$ the mean values of the pixels. If an edge E_{ij} is not included in any fault trace, $E_{ij} = 0$, the difference between the mean values of P_i and P_j should be $\mu_i - \mu_j = 0$. If there is a fault trace including the edge, $E_{ij} = 1$, the difference should be approximately $\mu_i - \mu_j = \bar{v}_{ij}$. Generally the requirement is thus

$$\mu_j - \mu_i = e_{ij}\bar{v}_{ij} \quad \forall (i \sim j), \tag{11}$$

where $i \sim j$ denotes adjacent pixels. The equation system is in general singular and has no solution. To overcome this problem the offset values of the vertices are not considered to give the final offsets along the fault traces. Instead the surface of mean values $\mu(\boldsymbol{e}, \boldsymbol{v})$ is approximated according to some optimization criterion. The pixel values only give relative heights of the surface, so without loss of generality the mean value of pixel 1 can be assigned a value $\mu_1 = 0$. The remaining mean values $\mu_2, \ldots, \mu_{n_p}$ are found as a least square solution of the linear equation system (11). The equation system can be expressed as a regression problem $\boldsymbol{Y}(\boldsymbol{e}, \boldsymbol{v}) = \boldsymbol{Z}\beta$ where the components of $\boldsymbol{Y}(\boldsymbol{e}, \boldsymbol{v})$ are given as $e_{ij}\bar{v}_{ij}$, and $\beta = (\mu_2, \mu_3, \ldots, \mu_{n_p})$. The design matrix $\boldsymbol{Z}$ is independent of $\boldsymbol{e}$ and $\boldsymbol{v}$, and has component values $Z_{ij} \in \{-1, 0, 1\}$. The least square fit $\hat{\beta}$ is given by $\hat{\beta} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}(\boldsymbol{e}, \boldsymbol{v})$, and this fit, together with $\mu_1 = 0$, is used as mean values of the pixels $\boldsymbol{P}$. Thus

$$\mu(\boldsymbol{e}, \boldsymbol{v}) = (0, (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}(\boldsymbol{e}, \boldsymbol{v})). \tag{12}$$

The entries of the covariance matrix Σ_p is given as

$$\mathrm{Cov}(P_i, P_j) = \sigma_p^2 \rho(x_i, x_j), \tag{13}$$

where x_i denotes the position of pixel i. The exponential correlation function is used,

$$\rho(x, x') = \rho(x' - x) = \exp\{-3(|x' - x|/a)^\nu\}, \qquad a > 0,\ 0 < \nu \leq 2, \tag{14}$$

where the parameter a is called the correlation range.

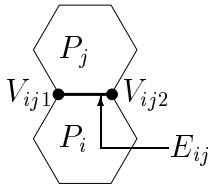


Figure 12: The figure illustrates the notation of adjacent pixels P_i and P_j, the edge E_{ij} between the pixels, and the vertices V_{ij1} and V_{ij2} at the ends of the edge.

3.2.4 Faulted horizon

The image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$ gives a discrete parameterization of the topography of a faulted horizon. A continuous representation H is created by triangulation between the pixel

centers, as described in Section 3.1. The topographic surface H is a deterministic function of the image $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$, and the pdf of H is given in Expression (2). Inserting Expressions (4), (8) and (10), the following expression of the pdf of H is obtained:

$$\begin{aligned} f(h) &= \text{const} \times \phi(\boldsymbol{p}; \mu(\boldsymbol{e}, \boldsymbol{v}), \Sigma_p) \cdot n_b(\boldsymbol{e})^{-1} f(\boldsymbol{v}_{b_0}) \prod_{b \in \mathcal{B}_e} f(\boldsymbol{v}_b) \cdot \exp\{-\sum_{c \in \mathcal{C}^m} \omega_c(\boldsymbol{e}_c)\} \\ &= \text{const} \times \exp\{-(\boldsymbol{p} - \mu(\boldsymbol{e}, \boldsymbol{v}))' \Sigma_p^{-1} (\boldsymbol{p} - \mu(\boldsymbol{e}, \boldsymbol{v}))/2 - \sum_{c \in \mathcal{C}^m} \omega_c(\boldsymbol{e}_c)\} \\ &\quad \times n_b(\boldsymbol{e})^{-1} \theta^k (1-\theta)^{|\mathcal{B}_e|-k}, \end{aligned} \tag{15}$$

where k is the number of faults $b \in \mathcal{B}_e$ offset in direction $D_b = 1$. The number $n_b(\boldsymbol{e})^{-1}$, the dimension $|\mathcal{B}_e|$ of the set $\mathcal{B}_e$, the sets b contained in $\mathcal{B}_e$ and the set b_0 depend on the edge values $\boldsymbol{E}$. Thus the number of terms in the product $\prod_{b \in \mathcal{B}_e} f(\boldsymbol{v}_b)$ will vary, and also the dimensions of the variables $\boldsymbol{V}_b$.

Figure 13 shows a realization of h from the prior pdf $f(h)$. The sampling procedure used to generate this realizations is described in Section 6. The figure shows edges $\boldsymbol{e}$, Figure 13a, vertices $\boldsymbol{v}$, Figure 13b, and pixels $\boldsymbol{p}$, Figure 13c. Figure 13d gives a perspective plot of the continuous surface $h = h(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{p})$, and the surface is represented as a gray scale image in Figure 13f. Figure 13e shows the final offsets, being the magnitudes of the discontinuities along fault traces in the surface. The figure should be compared with Figure 13b. Away from fault intersections the two figures appear to be quite similar, with only minor differences. Close to fault trace intersections the two figures differ to some extent, indicating that the mean values $\mathrm{E}[\boldsymbol{P}|\boldsymbol{E}, \boldsymbol{V}] = \mu(\boldsymbol{e}, \boldsymbol{v})$, Expression (12), are not able to match the vertex values exactly in these regions.

4 Likelihood Model of Reservoir Specific Observations

Faults alter the geometric structure of a reservoir, and thus influence on the structure of reservoir characteristics. Available observations of the characteristics of subsurface reservoirs are discussed in Section 4.1. The relationship between the observations and the fault pattern is defined through the likelihood function, and is described in Section 4.2.

4.1 Observations

Available observations from a subsurface reservoir consist of three dimensional seismic data and well observations. Observations from seismic surveys give a good spatial covering of the reservoir. However, only faults with displacement above 15–20 meters can be resolved

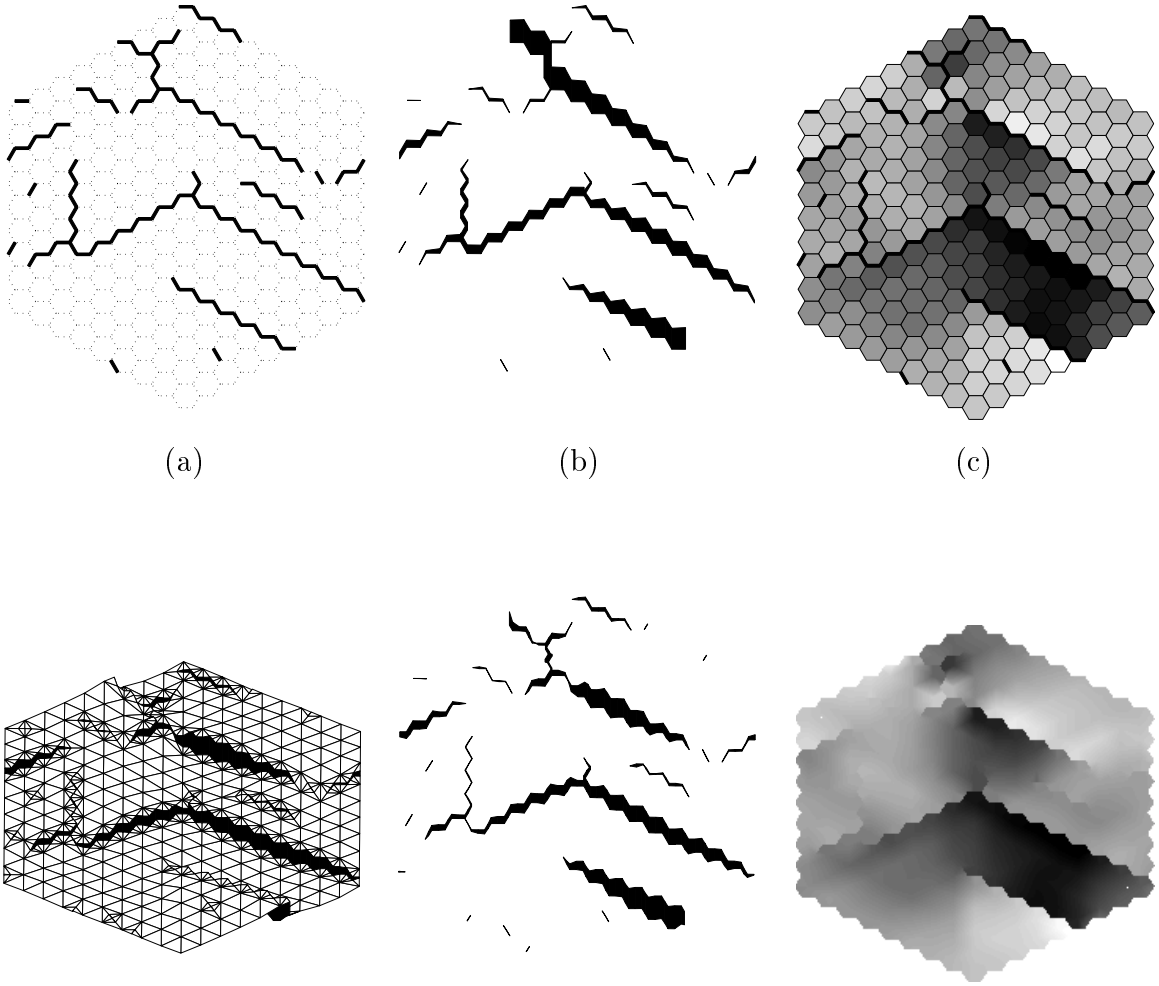


Figure 13: Realization h from the prior pdf (15). (a) Edges $\boldsymbol{e}$ from $f(\boldsymbol{e})$, (b) vertices $\boldsymbol{v}$ from $f(\boldsymbol{v}|\boldsymbol{e})$ and (c) pixels $\boldsymbol{p}$ from $f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v})$. (d) Perspective plot of the faulted horizon h, obtained by triangulation of the image $(\boldsymbol{e},\boldsymbol{v},\boldsymbol{p})$. (e) Fault planes from figure (d), showing the true offsets of the faults. (f) Horizon h represented as a gray-scale image.

from seismic data, and the observations contain measurement errors. See for example Sheriff and Geldart (1995) for an introduction to seismic theory. Fine scale observations of reservoir characteristics can be obtained from well logs. Such observations are however few and sparse. Under the assumption of vertical faults, macroscopic faults are not observable in wells. Seismic travel time is used as the vertical scale, while the conversion from time to depth is beyond the scope of this work.

4.1.1 Seismic data

When a seismic wave travels downwards and hits an acoustic impedance discontinuity, a part of the wave is reflected, and the reflected signal is registered at the surface. For P-waves, also denoted longitudinal or compressional waves, the acoustic impedance Z is given as the product of the P-wave velocity α and the density ρ of the medium, $Z = \alpha\rho$. Discontinuities in acoustic impedance are called reflectors, and the most common reflectors are unconformities and significant changes in lithology, see Sheriff and Geldart (1995). Consider a reflector where the P-wave velocity, density and acoustic impedance on one side is α_1, ρ_1 and $Z_1 = \alpha_1\rho_1$ respectively, and on the opposite side α_2, ρ_2 and $Z_2 = \alpha_2\rho_2$. The reflection coefficient at the discontinuity is defined as

$$C = \frac{Z_2 - Z_1}{Z_2 + Z_1}. \tag{16}$$

For a fixed point (x, y) in a horizontal grid of observations, the seismic amplitudes obtained from reflection points along a vertical axis in (x, y) are denoted $S^o(x, y, t)$. The seismic data can be expressed as a convolution of the reflection coefficients $C(x, y, t)$ with a seismic wavelet $w(t) = w(t; \psi_s)$, where ψ_s is a wavelet parameter. The scale of the vertical axis is travel time t, and the superscript o denotes observations throughout this work. A simplification is used, assuming all ray-paths of the seismic wave are vertical, and that all reflectors are horizontal. The seismic data are given by

$$S^o(x, y, t) = w(t) * C(x, y, t) + U(x, y, t) = \int_{-\infty}^{\infty} w(\tau) C(x, y, t - \tau)\, d\tau + U(x, y, t), \tag{17}$$

where $U(x, y, t) \sim N(0, \sigma_s^2)$ is a random noise term including both modeling error and measurement error, see Eide (1999). An approximation to (17) is given by a discrete representation. Assume seismic observations are available for times $t_1, t_2, \ldots, t_n$, and let $S_k^o = S^o(x, y, t_k)$ be the seismic observation at the point (x, y, t_k). The discrete representation of the seismic observation is as follows:

$$S_k^o = \sum_{i=-L}^{L} w_i c_{k-i} \delta t + U_k, \tag{18}$$

where $U_k = U(x, y, t_k)$ is the noise term, $w_i = w(t_i)$ is the seismic wavelet evaluated at time t_i, $c_{k-i} = C(x, y, t_{k-i})$ is the reflection coefficient at time t_{k-i} and $\delta t = t_{i+1} - t_i$ is

the time step. Corresponding to Expression (18), an expression of all seismic observations, denoted $\boldsymbol{S}^o$, is given as

$$\boldsymbol{S}^o = A(\psi_s)\boldsymbol{C} + \boldsymbol{U}, \tag{19}$$

where $\boldsymbol{C}$ is a matrix of reflection coefficients, $A(\psi_s)$ is a matrix with entries corresponding to the product $w_i \delta t$ and $\boldsymbol{U}$ a matrix of independent random noise. A commonly used wavelet is the Ricker wavelet, expressed as

$$w(t) = (1 - 2(\pi \nu_M t)^2) \exp(-(\pi \nu_M t)^2), \tag{20}$$

where the wavelet parameter $\psi_s = \nu_M$ is the peak frequency, see Sheriff and Geldart (1995). The Ricker wavelet is symmetric with a maximum at $t = 0$, see Figure 14.

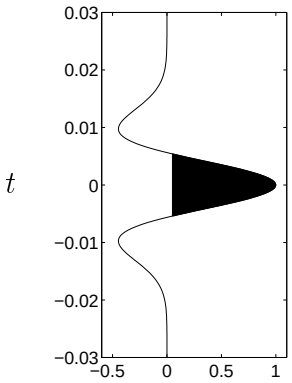


Figure 14: The Ricker wavelet (20) with peak frequency $\nu_M = 40Hz$.

4.1.2 Well observations

From well logs velocity α and density ρ can be derived. The acoustic impedance $Z = \alpha\rho$ can then be calculated, and for vertical wells Expression (16) can be used to obtain reflection coefficients $C(x, y, t)$ in a well located at (x, y). Denote the vector of observed reflection coefficients $\boldsymbol{C}^o$. The observations of reflection coefficients from wells are a subset of the reflection coefficients $\boldsymbol{C}$. Throughout this work well observations are assumed to be exact observations.

4.2 Likelihood model

Two types of observations are discussed in Section 4.1, seismic data and observations of reflection coefficients from well logs. The likelihood model of the seismic data and

reflection coefficients is based on the work by Eide (1999). See also Eide et al. (1997a), Eide et al. (1997b). The likelihood function of the reflection coefficients $\boldsymbol{C}$ and $\boldsymbol{C}^o$ is discussed first. The likelihood function of the seismic data $\boldsymbol{S}^o$ and the joint likelihood function of all observations $\boldsymbol{O} = (\boldsymbol{S}^o, \boldsymbol{C}^o)$ are obtained from Expression (17) or (19), based on the likelihood function of the reflection coefficients.

4.2.1 Reflection coefficients

A likelihood function of reflection coefficients in a non-faulted reservoir is first defined. Adjustments are then made to find the likelihood function of the reflection coefficients in the faulted reservoir.

Consider a sedimentary reservoir prior to faulting. Due to the sedimentation process the rock tends to have a layered structure, with a relatively strong lateral homogeneity. Vertically the layered rock is non-homogeneous. Prior to faulting, reflection coefficients of a layered rock can in some cases be modeled as a Gaussian random field, see Todoeschuck et al. (1990) and Eide (1999). Let $\boldsymbol{C}_u$ denote reflection coefficients of a non-faulted reservoir, with a Gaussian distribution

$$\boldsymbol{C}_u \sim N(\mu_{c_u}, \Sigma_{c_u}), \tag{21}$$

where μ_{c_u} is a vector of mean values and Σ_{c_u} is the covariance matrix of $\boldsymbol{C}_u$. The covariance between reflection coefficients in the points (x_i, y_j, t_k) and (x_l, y_m, t_n) is given as

$$\mathrm{Cov}\{C_u(x_i, y_j, t_k), C_u(x_l, y_m, t_n)\} = \sigma^2_{c_u} \rho_{c_u}((x_i, y_j, t_k), (x_l, y_m, t_n)), \tag{22}$$

where $\rho_{c_u}(\cdot)$ is the correlation function of the reflection coefficients and $\sigma^2_{c_u}$ the variance. The lateral homogeneity of the rock is accounted for by defining a strong horizontal correlation in the likelihood model. A correlation function with negative correlation at short vertical ranges is used, suggested by Eide (1999). The horizontal and vertical correlations are assumed to be separable, and are defined based on the exponential correlation function (14). Let $\rho_H(\cdot)$ and $\rho_V(\cdot)$ denote exponential correlation function with range $a = a_H$ and $a = a_V$ respectively. The correlation between $C_u(x_i, y_j, t_k)$ and $C_u(x_l, y_m, t_n)$ can be expressed as

$$\begin{aligned} \rho_{c_u}((x_i, y_j, t_k), (x_l, y_m, t_n)) \;&=\; \rho_H((x_i, y_j), (x_l, y_m)) \times \\ &\{2(1 - \rho_V(\delta t))\}^{-1}\{2\rho_V(t_k, t_n) - \rho_V(t_{k+1}, t_n) - \rho_V(t_k, t_{n+1})\} \end{aligned} \tag{23}$$

where $\rho_H((x_i, y_j), (x_l, y_m))$ is the horizontal correlation and the expression on the last line is the vertical correlation. A regular grid with vertical spacing $t_{i+1} - t_i = \delta t$ is used. A strong horizontal correlation is obtained by assigning a large value to the correlation range a_H. The vertical correlation, corresponding to $(i, j) = (l, m)$ in the correlation function (23), is illustrated in Figure 15.

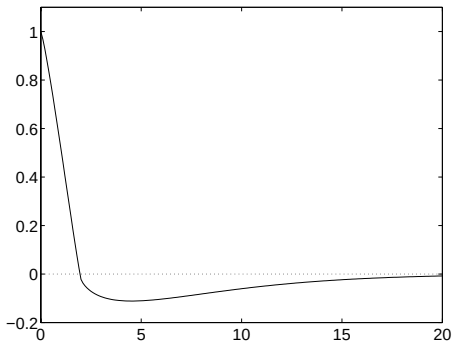


Figure 15: The vertical correlation in the correlation function (23), with parameters $\nu = 1.2$ and $a_V = 15$ in the exponential correlation function $\rho_V(\cdot)$ involved in Expression (23). The scale of the horizontal axis is milliseconds.

After faulting the original structure of the rock is altered. The fault pattern is represented by the horizon $H(x, y)$. Under the basic model assumptions discussed in Section 2, the reflection coefficients originally located in (x, y, t) is displaced to the point $(x, y, t-H(x, y))$, see Figures 16a and 16b. If the faulted horizon $H(x, y)$ is known, the reflection coefficients can be transformed back to their original location, as illustrated in Figure 16c.

Let $C(x, y, t)$ denote the reflection coefficient at a point (x, y, t) in the faulted reservoir. The value of this reflection coefficient is

$$C(x, y, t) = C_u(x, y, t + H(x, y)), \tag{24}$$

obtained by a vertical transformation of the original reflection coefficients C_u. The marginal distributions of the reflection coefficients are unchanged, but due to the spatial rearrangement the mean vector μ_{c_u} and covariance matrix Σ_{c_u} are no longer valid. Consider two points (x, y, t) and (x', y', t') prior to faulting, with reflection coefficients C_u and C'_u. The covariance between the reflection coefficients prior to faulting is given in Expression (22).

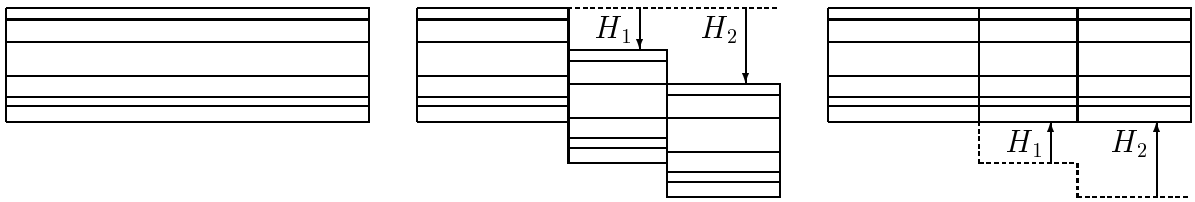


Figure 16: (a) The structure of the rock before faulting. (b) After faulting the structure is altered. (c) If the faulted horizon H is known, a transformation back to the original structure can be performed.

After faulting the covariance between the two reflection coefficients is unchanged, since the values of the reflection coefficients are not altered by the faulting. However, the positions are changed, and the reflection coefficients considered are now located at $(x, y, t - H(x, y))$ and $(x', y', t' - H(x', y'))$ respectively. The transformations due to faulting can be expressed as a permutation of the elements of the vector $\boldsymbol{C}_u$ of reflection coefficients, and a corresponding rearrangement of the order of the elements in the mean vector μ_{c_u} and covariance matrix Σ_{c_u}. The overall distribution of reflection coefficients in a faulted reservoir is then a Gaussian distribution, expressed as

$$[\boldsymbol{C}|H = h] \sim N(\mu_c(h), \Sigma_c(h)), \tag{25}$$

where $\mu_c(h)$ and $\Sigma_c(h)$ are obtained from μ_{c_u} and Σ_{c_u} by rearranging the positions of the elements according to the transformations dictated by the faulted horizon $H = h$. Reflection coefficients $\boldsymbol{C}^o$ observed in wells are a subset of the reflection coefficients $\boldsymbol{C}$. The observed reflection coefficients follow a Gaussian distribution

$$[\boldsymbol{C}^o|H = h] \sim N(\mu_{c^o}(h), \Sigma_{c^o}(h)), \tag{26}$$

where $\mu_{c^o}(h)$ and $\Sigma_{c^o}(h)$ are subsets of $\mu_c(h)$ and $\Sigma_c(h)$ respectively.

4.2.2 Seismic data

The likelihood function of seismic data $\boldsymbol{S}^o$ is obtained by using the relationship between seismic observations and reflection coefficients given in Expression (17) or (19), and the likelihood model of the reflection coefficients given in Expression (25). The seismic data are linear combinations of the reflection coefficients and a random noise term, both Gaussian random fields, thus the seismic observations are also modeled as a Gaussian random field. The mean and variance functions of the seismic data are found from the mean and variance functions of the reflection coefficients and the random noise term.

Similar to the reflection coefficients, the seismic data can be expressed as a transformation:

$$S^o(x, y, t) = S^o_u(x, y, t + H(x, y)). \tag{27}$$

Using the same notation as in Expression (24), S^o_u denotes the seismic observation that would be obtained if seismic data were gathered prior to faulting of the reservoir. The mean value of $S^o_u(x, y, t)$ is found from Expression (17):

$$\mathrm{E}\{S^o_u(x, y, t)\} = \int_{-\infty}^{\infty} w(\tau)\mu_{c_u}(x, y, t - \tau)\, d\tau. \tag{28}$$

The covariance between seismic observations $S^o_u(x_i, y_j, t_k)$ and $S^o_u(x_l, y_m, t_n)$ is found to be

$$\begin{aligned}\mathrm{Cov}\{S^o_u(x_i, y_j, t_k), S^o_u(x_l, y_m, t_n)\} \;&=\; \sigma_s^2 \mathbf{1}[(i, j, k) = (l, m, n)] + \\ \sigma^2_{c_u} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} &w(\tau_1)w(\tau_2)\rho_{c_u}((x_i, y_j, t_k - \tau_1), (x_l, y_m, t_n - \tau_2))d\tau_1\, d\tau_2\end{aligned} \tag{29}$$

where $\mathbf{1}[\cdot]$ is the indicator function. The covariance function (22) and the correlation function (23) of the reflection coefficients are used. Using the discrete matrix representation (19), the mean values and covariance matrix of the seismic data $\boldsymbol{S}_u^o$ are found to be

$$\mu_{s_u^o} = \mathrm{E}\{\boldsymbol{S}_u^o\} = A(\psi_s)\mu_{c_u} \qquad \text{and} \qquad \Sigma_{s_u^o} = \mathrm{Var}\{\boldsymbol{S}_u^o\} = A(\psi_s)\Sigma_{c_u}A(\psi_s)' + \sigma_s^2 I, \tag{30}$$

where I is the identity matrix. As for the reflection coefficients, the seismic observations of the faulted reservoir can be modeled as a Gaussian random field with mean vector and covariance matrix depending on the faulted horizon H. The distribution of the seismic data is the Gaussian distribution

$$[\boldsymbol{S}^o|H = h] \sim N(\mu_{s^o}(h), \Sigma_{s^o}(h)) \tag{31}$$

where $\mu_{s^o}(h)$ and $\Sigma_{s^o}(h)$ are obtained by rearranging the elements of $\mu_{s_u^o}$ and $\Sigma_{s_u^o}$ according to the faulted horizon $H = h$.

4.2.3 Joint likelihood model of observations

Reflection coefficients $[\boldsymbol{C}|H]$ are Gaussian distributed, see Expression (25). From Expression (19) it is clear that the joint distribution of $[(\boldsymbol{S}^o, \boldsymbol{C})|H]$ is also Gaussian, and consequently, so is the distribution of the subset $[\boldsymbol{O}|H]$ of observations, where $\boldsymbol{O} = (\boldsymbol{S}^o, \boldsymbol{C}^o)$. By the same arguments as above, the mean vector and covariance matrix of $[\boldsymbol{O}|H]$ can be found by first deriving the corresponding mean and covariance in a non-faulted reservoir, using Expressions (17) or (19), and then rearranging the elements according to the faulted horizon $H = h$. Thus $[\boldsymbol{O}|H = h] \sim N(\mu_o(h), \Sigma_o(h))$, and the joint likelihood function $f(\boldsymbol{o}|h) = f(\boldsymbol{s}^o, \boldsymbol{c}^o|h)$ of the observations $\boldsymbol{O} = (\boldsymbol{S}^o, \boldsymbol{C}^o)$, given H, is the pdf of the Gaussian distribution:

$$f(\boldsymbol{o}|h) = \phi(\boldsymbol{o}; \mu_o(h), \Sigma_o(h)), \tag{32}$$

see Expression (10). The single elements of the covariance matrix $\Sigma_o(h)$ are defined by the covariance function (22) and the matrix $A(\psi_s)$ involved in the convolution (19), while the order of the elements are determined by the faulted horizon H.

5 Posterior Model of Geological Fault Patterns Conditioned to Reservoir Specific Observations

Observations $\boldsymbol{O}$ of the reservoir contain seismic data and observations of reflection coefficients obtained from well logs. The observations carry information about the structure of the reservoir, and how the originally horizontal layers of the reservoir have been altered due to faulting. The aim of this work is to generate realizations of the underlying geological

fault pattern, by sampling the faulted horizon H conditioned to the observations. The prior distribution from Section 3 and the likelihood function from Section 4 are combined to form a posterior distribution. Samples from the posterior distribution can be obtained using the methodology presented in Section 6.

The prior pdf $f(h)$ contains general geological knowledge about the fault pattern and the faulted horizon H. The likelihood function $f(\boldsymbol{o}|h)$ gives the likelihood of observing $\boldsymbol{O}$, if the true faulted horizon is H. The posterior distribution

$$f(h|\boldsymbol{o}) = \text{const} \times f(h) f(\boldsymbol{o}|h) \tag{33}$$

combines the general geological knowledge with the reservoir specific observations, and gives the distribution of the faulted horizon H conditioned to the observations $\boldsymbol{O}$. The prior pdf and the likelihood function are given in Expressions (15) and (32) respectively. Inserting the functions into Expression (33), the following posterior pdf is obtained:

$$\begin{aligned} f(h|\boldsymbol{o}) \;\; = \;\; & \text{const} \times \phi(\boldsymbol{p}; \mu(\boldsymbol{e}, \boldsymbol{v}), \Sigma_p) \cdot n_b(\boldsymbol{e})^{-1} f(\boldsymbol{v}_{b_0}) \prod_{b \in \mathcal{B}_e} f(\boldsymbol{v}_b) \cdot \exp\{-\sum_{c \in \mathcal{C}^m} \omega_c(\boldsymbol{e}_c)\} \\ & \times \phi(\boldsymbol{o}; \mu_o(h), \Sigma_o(h)). \end{aligned} \tag{34}$$

The posterior model contains a number of parameters, which are listed in Table 4.

	Prior distribution $f(h) = f(\boldsymbol{p}\|\boldsymbol{e}, \boldsymbol{v}) f(\boldsymbol{v}\|\boldsymbol{e}) f(\boldsymbol{e})$*:*
$f(\boldsymbol{e})$:	Potential function $\omega_c(\boldsymbol{e}_c)$: γ_i, $\gamma_{i,j}$, $i = 1, \ldots, 10$, $j = 1, \ldots, 6$
$f(\boldsymbol{v}\|\boldsymbol{e})$:	Dip parameter θ
	Size parameter μ_v
$f(\boldsymbol{p}\|\boldsymbol{e}, \boldsymbol{v})$:	Variance σ_p^2
	Correlation function $\rho(\cdot)$: ν, a

	Likelihood function $f(\boldsymbol{o}\|h) = f(\boldsymbol{s}^o\|\boldsymbol{c}^o, h) f(\boldsymbol{c}^o\|h)$*:*
$f(\boldsymbol{c}^o\|h)$:	Mean μ_c
	Variance σ_c^2
	Correlation function $\rho_c(\cdot)$: ν, a_H, a_V
$f(\boldsymbol{s}^o\|\boldsymbol{c}^o, h)$:	Variance σ_s^2
	Ricker wavelet $w(t)$: ν_M
	Time step δt

Table 4: Parameters in the posterior distribution $f(h|\boldsymbol{o})$.

6 Sampling From the Posterior Distribution

Expression (34) gives the posterior pdf of the faulted horizon $H(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$, conditioned to the observations $\boldsymbol{O}$. Fault patterns and corresponding faulted horizons can be generated by sampling from the posterior distribution. The posterior distribution is high dimensional, and due to the complexity of the distribution no procedure for exact simulation exists. Instead, Markov chain Monte Carlo (McMC) techniques are used, see Besag et al. (1995) for an introduction to McMC methodology and for further references.

A Metropolis-Hastings algorithm is used to generate samples from the posterior distribution $f(h|\boldsymbol{o})$ in (34), see Hastings (1970). A detailed description is given in Algorithm 1, Appendix A. In each step of the algorithm, one of three possible transition types is chosen. Suppose at one step the edge, vertex and pixel values are $\boldsymbol{E} = \boldsymbol{e}$, $\boldsymbol{V} = \boldsymbol{v}$ and $\boldsymbol{P} = \boldsymbol{p}$, and the faulted horizon is $H(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P}) = h(\boldsymbol{e}, \boldsymbol{v}, \boldsymbol{p})$. New values $\boldsymbol{e}'$, $\boldsymbol{v}'$ and $\boldsymbol{p}'$ are suggested using one of the transition types. The three alternative transition types have different transition kernels, and are given as follows:

(i) Edge values $\boldsymbol{e}'$, vertex values $\boldsymbol{v}'$ and pixel values $\boldsymbol{p}'$ are generated from the prior distribution, with the transition kernel

$$q(h'|h) = f(\boldsymbol{e}', \boldsymbol{v}', \boldsymbol{p}') = f(\boldsymbol{e}')f(\boldsymbol{v}'|\boldsymbol{e}')f(\boldsymbol{p}'|\boldsymbol{e}', \boldsymbol{v}').$$

(ii) Edge values $\boldsymbol{e}' = \boldsymbol{e}$ are fixed while vertex values $\boldsymbol{v}'$ and pixel values $\boldsymbol{p}'$ are generated from their respective prior distributions. The transition kernel is in this case

$$q(h'|h) = f(\boldsymbol{v}', \boldsymbol{p}'|\boldsymbol{e}) = f(\boldsymbol{v}'|\boldsymbol{e})f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}').$$

(iii) Edge values $\boldsymbol{e}' = \boldsymbol{e}$ and vertex values $\boldsymbol{v}' = \boldsymbol{v}$ are fixed while pixel values $\boldsymbol{p}'$ are generated from the prior distribution. In this case the transition kernel is

$$q(h'|h) = f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}).$$

In alternative (i) and (ii) the prior density function $f(\boldsymbol{v}|\boldsymbol{e})$ cancels in the acceptance probability, while in alternative (iii) the vertex values $\boldsymbol{V}$ are constant. Thus the dependence of $\mathcal{B}_e$ on $\boldsymbol{E}$ causes no problems in the McMC algorithm, and the number $n_b(\boldsymbol{e})^{-1}$ need not be evaluated. For all suggested transitions, the probability of accepting the proposed horizon $h' = h'(\boldsymbol{e}', \boldsymbol{v}', \boldsymbol{p}')$ is given by

$$\alpha(h'|h) = \min\left\{1,\ \frac{f(\boldsymbol{o}|h')}{f(\boldsymbol{o}|h)}\right\}. \tag{35}$$

A sample from the pdf $f(h)$ is generated in a stepwise procedure, and Figure 13 illustrates the realizations obtained at different steps of the procedure. First, all edge values $\boldsymbol{E}$ are generated from the prior pdf $f(\boldsymbol{e})$. Next, vertex values $\boldsymbol{V}$ are generated from the prior pdf

$f(\boldsymbol{v}|\boldsymbol{e})$ conditioned on the edge values. The pixel values $\boldsymbol{P}$ are then generated from the prior pdf $f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v})$ conditioned on both edge and vertex values.

A Metropolis-Hastings algorithm is used to generate realizations from the prior pdf $f(\boldsymbol{e})$, while exact sampling of the offsets and pixels from $f(\boldsymbol{v}|\boldsymbol{e})$ and $f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v})$ are easily performed. Algorithms for drawing samples of $\boldsymbol{E}$, $\boldsymbol{V}$ and $\boldsymbol{P}$ from their respective prior distributions are given in Appendix A.

The sample space $\Omega = \Omega_e^{n_e} \times \Omega_v^{n_v} \times \Omega_p^{n_p}$ of H is a high dimensional space, and a large number of iterations must be run in order to span the space adequately. Thus the algorithm is rather time consuming and as n_e, n_v and n_p increase, the number of iterations required increases drastically. In alternative (i) of the transition steps, one further McMC simulation is run within a step of the original McMC algorithm. Thus realizations from the posterior distribution is generated using a double McMC algorithm. However, the McMC algorithm used to sample from the prior pdf $f(\boldsymbol{e})$ is not very time consuming and seems to converge relatively fast, see Section 7.

The simplicity of the likelihood function is an advantage of the described algorithm. All acceptance probabilities are given as a likelihood ratio, where the likelihood function (32) is the pdf of a multivariate Gaussian distribution. Thus the algorithm is capable of handling densely scattered seismic observations, although evaluation of the likelihood function in the case of a large number of observations can be time consuming. Still, the major influence on the time consumption of the algorithm is from the large number of iterations required to span the sample space of H through generation of samples from the prior pdf.

For small datasets the Gaussian likelihood function can be evaluated exactly. As the dimension increases however, the likelihood can in practice not be obtained analytically for general Gaussian fields, and an approximation to the likelihood function must be found. The likelihood function of the observations $\boldsymbol{O}$ can be approximated by a pseudo-likelihood function (Besag, 1974)

$$f(\boldsymbol{o}|h) \approx \prod_i f(\boldsymbol{o}_i|\boldsymbol{o}_{-i},h) \approx \prod_i f(\boldsymbol{o}_i|\boldsymbol{o}_{\partial i},h), \tag{36}$$

where $\boldsymbol{O}_i$ is chosen as a vertical column through the observations and $\boldsymbol{O}_{-i}$ is the set of all other observations. The conditional distributions of $\boldsymbol{O}_i$ are approximated by distributions depending only on observations $\boldsymbol{O}_{\partial i}$ in a neighborhood ∂i of $\boldsymbol{O}_i$. A neighborhood of 5×5 vertical columns is used in the simulations presented below.

7 Sampling From the Prior Model of Fault Traces

The McMC algorithm presented in Section 6 includes sampling from the prior distribution of H, where McMC techniques are also applied to sample from $f(\boldsymbol{e})$. Realizations from

$f(\boldsymbol{e})$ are shown in Figures 9 and 10, and are generated using the McMC algorithm described in Appendix A, Algorithm 2. The convergence of this McMC algorithm is studied in this section.

Figure 17 shows convergence plots for realizations from the prior pdf $f(\boldsymbol{e})$, using the potentials from example 1 in Table 3. The McMC algorithm is run for 20 000 iterations. Each iteration consists of n_e update steps, where at each step an edge is drawn at random, a change of value is suggested, and the new value is accepted with the appropriate probability. The plots in Figure 17 show the relative number of configurations belonging to some of the configuration classes from Table 1, plotted for every 10th iteration. All plots indicate a fast convergence of the algorithm. The Markov chain in Figure 17 has an initial state where each edge value is drawn at random from $\Omega_e = \{0, 1\}$. In Figure 18 some of the plots are compared to Markov chains starting at extreme initial states, one where all edge values are $E = 0$ and one where all are $E = 1$. The first 1000 iterations from the McMC algorithm are plotted. Within these iterations, the Markov chains seem to have reached the same stationary state. Figure 19 shows convergence plots for realizations from the prior pdf $f(\boldsymbol{e})$, using the anisotropic potentials from Simulation 8. Configurations of the favored direction are observed to be more frequently represented in the Markov chain. Also Figure 19 gives an indication of fast convergence.

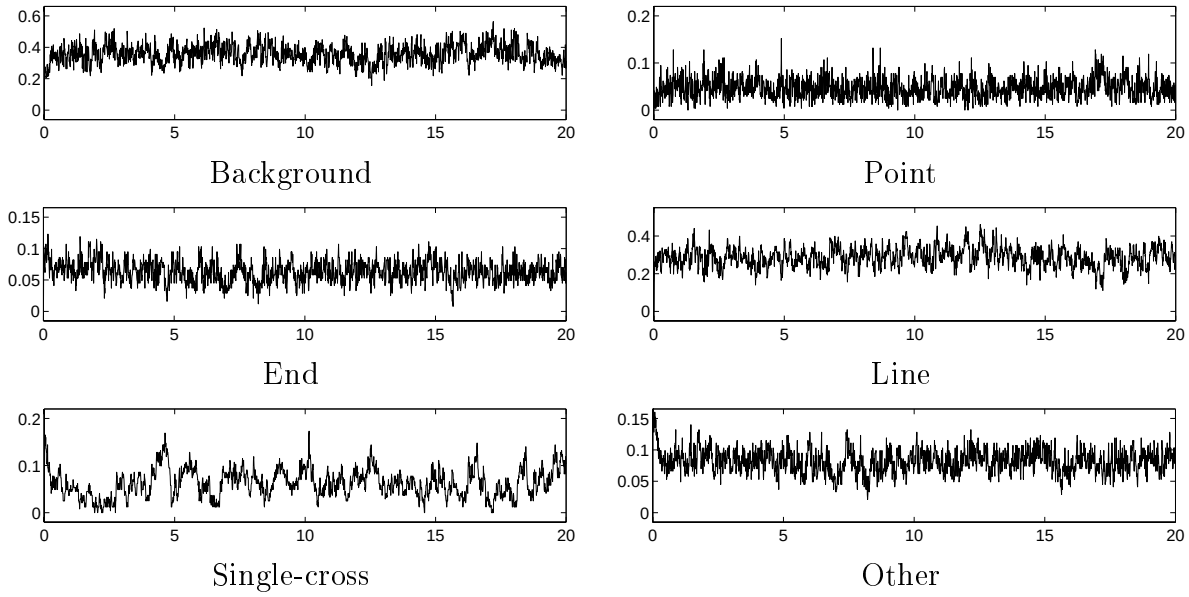


Figure 17: Plots of the relative number of configurations belonging to some of the classes listed in Table 1. The potentials of example 1 in Table 3 are used. The number of iterations is 20 000 and every 10th iteration is plotted. The scale of the horizontal axis is 1000.

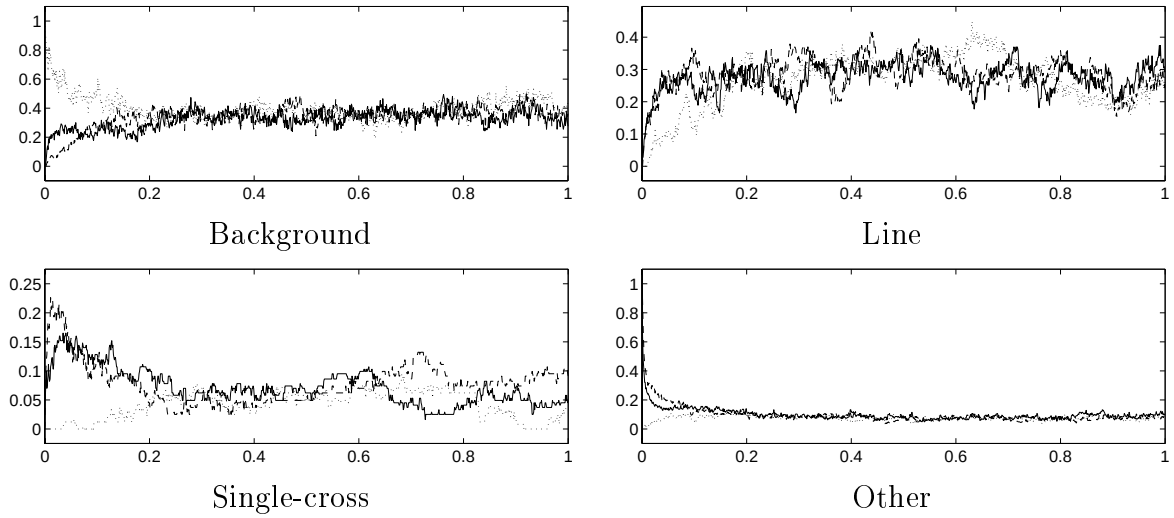


Figure 18: Plots of the relative number of configurations belonging to some of the classes listed in Table 1, using various initial states of the Markov chain. The potentials of example 1 in Table 3 are used. The first 1000 iterations from the prior pdf $f(\boldsymbol{e})$ are plotted, and the scale of the horizontal axis is 1000. Solid lines correspond to initial states where each edge value is drawn at random. Dotted lines show simulations where all edges have initial value $E = 0$, for dashed lines all initial values are $E = 1$.

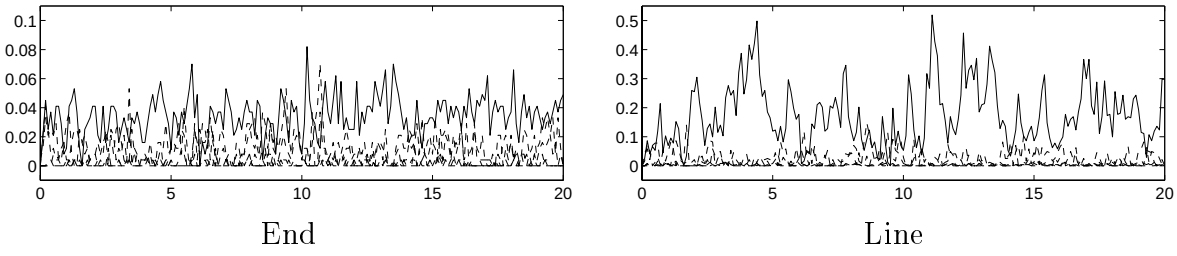


Figure 19: Plots of the relative number of configurations belonging to some of the classes listed in Table 2. The potentials of example 8 in Table 3 are used. The number of iterations is 20 000 and every 10th iteration is plotted. The scale of the horizontal axis is 1000. Solid lines represent configurations belonging to direction 2, the favored direction, while the other are plotted with dashed lines.

8 Simulation Examples using Synthetic Observations

The aim of this work is to generate faulted horizons H conditioned to observations $\boldsymbol{O}$, using the sampling technique described in Section 6. In this section synthetic data is used to study the model and the proposed algorithm. In Section 8.1 the synthetic data is presented, and in Section 8.2 simulation examples from the posterior distribution are studied. In the examples in Section 8.2 a small number of edges, vertices and pixels is chosen, to reduce the dimension of the sample space of the prior distribution.

8.1 Synthetic observations

To generate synthetic observations, data from a non-faulted reservoir is first generated. Reflection coefficients $\boldsymbol{C}_u \sim N(\mu_{c_u}, \Sigma_{c_u})$ of a non-faulted reservoir are drawn from a Gaussian distribution with mean μ_{c_u} and covariance matrix Σ_{c_u}, see Expression (21), using the covariance and correlation function given in Expressions (22) and (23) respectively. The size of the data set is $50 \times 50 \times 20$, and cross sections through the reflection coefficients are shown in Figure 20. The choice of parameter values are inspired by the work of Eide (1999), and are given in Table 5a. Seismic data $\boldsymbol{S}_u^o = A(\psi_s)\boldsymbol{C}_u + \boldsymbol{U}$ of the non-faulted reservoir are obtained by convolution of the reflection coefficients with the Ricker wavelet (20), adding Gaussian noise of mean 0 and variance σ_s^2, see Expression (19). The parameter ψ_s of the Ricker wavelet, the variance and the time step δt are given in Table 5b. Cross sections through the generated seismic data are shown in Figure 21. The generated reflection coefficients $\boldsymbol{C}_u$ and seismic data $\boldsymbol{S}_u^o$ represent the reservoir prior to faulting in the examples below.

A synthetic faulted horizon H is sampled from the prior pdf $f(h)$ in Expression (15), using Algorithm 1, Appendix A. The number of concentric bands of pixels in the hexagonal image is 5. The parameter values used in the prior pdf are given in Table 5c. The the horizon covers a hexagonal area, while the seismic observations are generated over a square area. The square area is placed inside the hexagonal area, and only the part of the horizon where observations are available is illustrated in Figure 22. Seismic data of the faulted reservoir are obtained using Expression (27), $S^o(x, y, t) = S_u^o(x, y, t + H(x, y))$, where $t + H(x, y)$ is rounded off to the closest observation point t_k. The faulted horizon in Figure 22 gives the seismic data shown in Figure 23. Reflection coefficients in four wells are found using Expression (24), $C^o(x, y, t) = C_u(x, y, t + H(x, y))$, and are shown in Figure 24. The well positions are picked at random, and are marked in Figure 23. The complete set of synthetic observations $\boldsymbol{O}$ consists of the seismic data $\boldsymbol{S}^o$ and the reflection coefficients $\boldsymbol{C}^o$.

When the seismic amplitudes $\boldsymbol{S}_u^o$ in Figure 21 are transformed according to Expression (27), columns of the images in Figure 21b-c are shifted up or down, depending on the values of $H(x, y)$. The result of the transformations is a non-rectangular area of observations. To

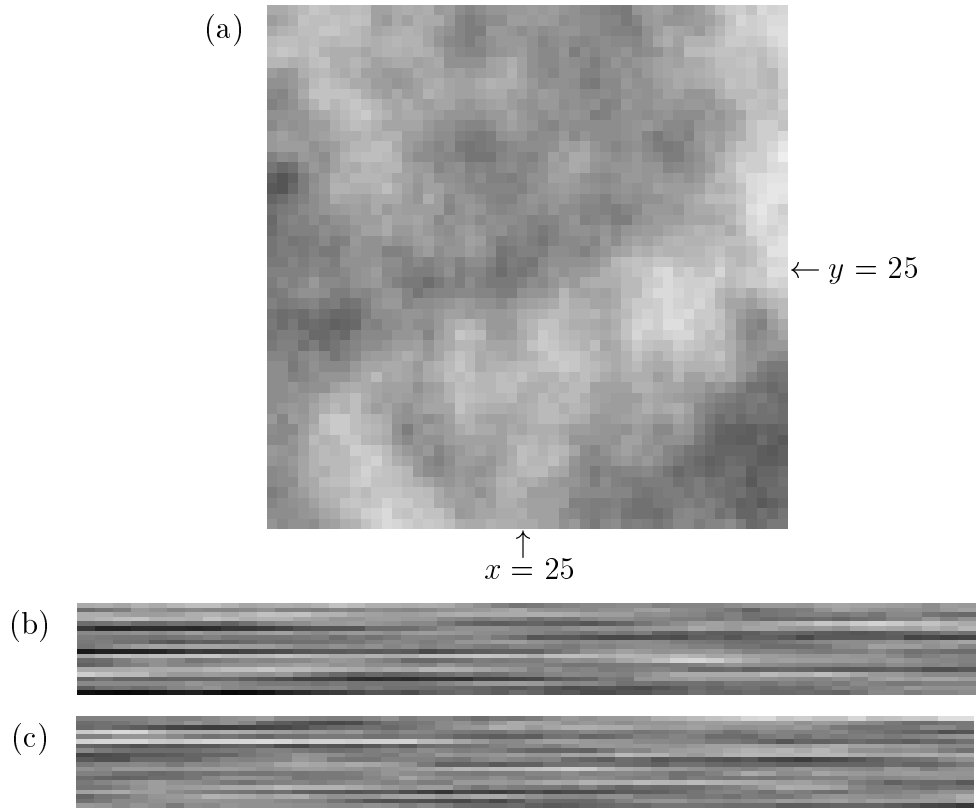


Figure 20: Synthetic reflection coefficients $\boldsymbol{C}_u$ in an unfaulted reservoir. (a) Top horizontal layer. (b) Vertical N-S oriented cross section through the reservoir at position $x = 25$. (c) E-W oriented cross section at $y = 25$. Different horizontal scales are used.

μ_c	σ_c^2	ν	a_H	a_V
0.0	0.035^2	1.2	40	15

(a)

σ_s^2	ν_M	δt
0.0015^2	40	0.002

(b)

$f(\boldsymbol{e})$									
γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
-1.0	0.7	-1.0	0.5	-1.1	0.125	-0.9	-0.8	0.2	2.5

$f(\boldsymbol{v}\|\boldsymbol{e})$	
μ_v	θ
0.01	0.5

$f(\boldsymbol{p}\|\boldsymbol{e},\boldsymbol{v})$		
σ_p^2	ν	a
0.01^2	2	3

(c)

Table 5: Parameters involved in (a) the Gaussian distribution of reflection coefficients $\boldsymbol{C}_u$ prior to faulting, (b) the relationship between reflection coefficients $\boldsymbol{C}_u$ and seismic data $\boldsymbol{S}_u^o$ and (c) the prior distribution $f(h) = f(\boldsymbol{e})f(\boldsymbol{v}|\boldsymbol{e})f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v})$ of the fault pattern H.

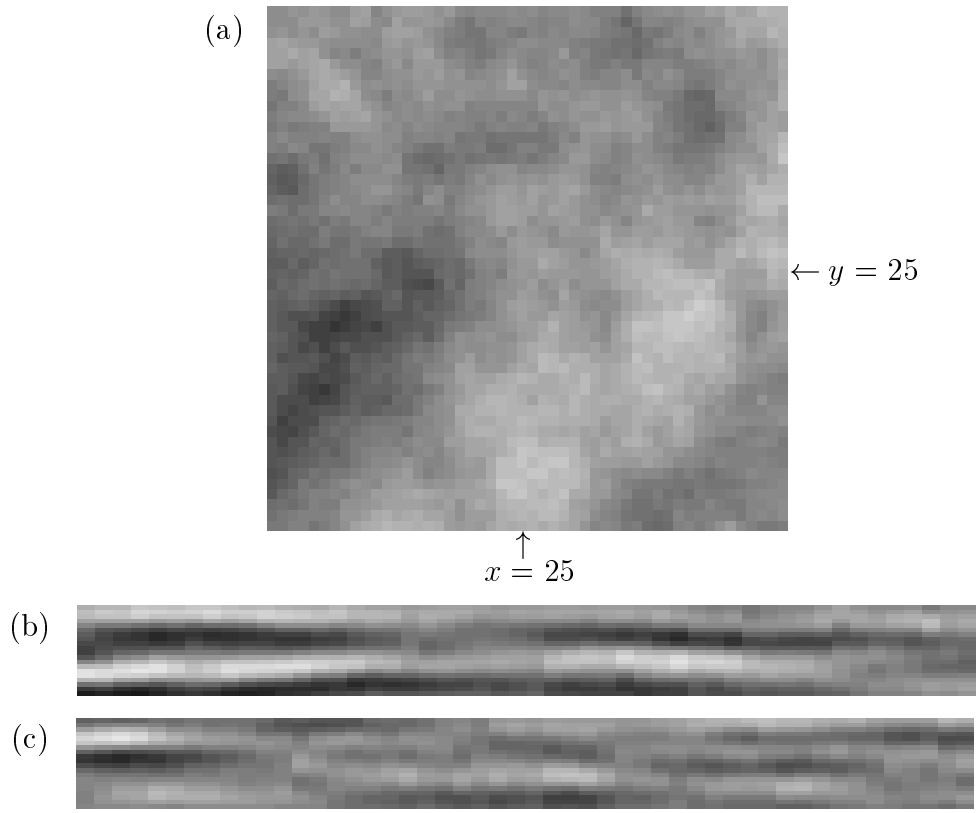


Figure 21: Seismic data $\boldsymbol{S}_u^o$ resulting from a convolution of the reflection coefficients in Figure 20 with the Ricker wavelet, and Gaussian noise added. (a) Top horizontal layer. (b) Vertical N-S oriented cross section through the reservoir at position $x = 25$. (c) E-W oriented cross section at $y = 25$. Different horizontal scales are used.

overcome this problem, only the 8 middle layers of the data set is used, and all shifts are restricted to having maximum size 6. Thus the 8 layers in the middle will always contain data after the transformations, and only these data are used as seismic observations $\boldsymbol{S}^o$. The size of the seismic dataset is thus $50 \times 50 \times 8$. Similar restrictions are used for the observed reflection coefficients from wells. Note that the observation points lying inside this rectangular area after transformation will differ according to the faulted horizon H. In visual representations of the seismic data or well data, all data are included, and the part of a data column that is transformed below the boundary of the image is added on the top, see Figure 23. Similar for data transformed above the image boundary.

Realizations H generated from the posterior distribution of the faulted horizon conditioned to synthetic observations can be compared to the true horizon. Furthermore, given the fault pattern H the observations can be transformed back to their original positions by inverting Expressions (24) and (27). For a given realization of H the transformed seismic data can be compared to the true non-faulted reservoir $\boldsymbol{S}_u^o$ illustrated in Figure 21.

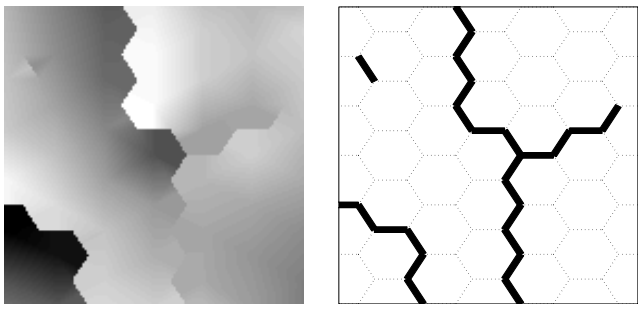


Figure 22: (a) Faulted horizon h from $f(h)$ and (b) corresponding pattern of fault traces $\boldsymbol{e}$.

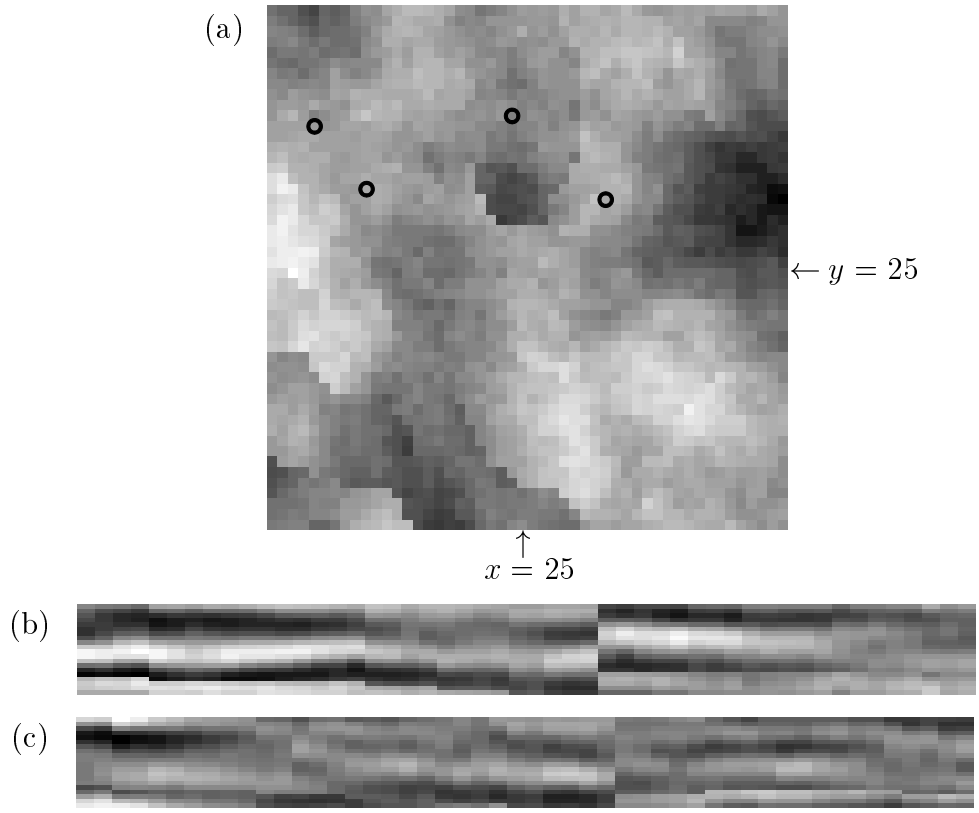


Figure 23: Seismic data $\boldsymbol{S}^o$ after faulting, corresponding to the data in Figure 21. (a) Top horizontal layer, where the circles mark the well positions. (b) Vertical N-S oriented cross section at position $x = 25$. (c) E-W oriented cross section at $y = 25$. Different horizontal scales are used.

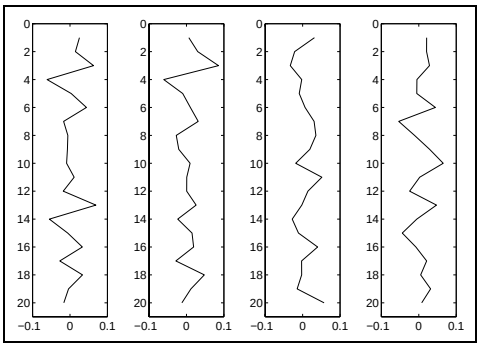


Figure 24: Reflection coefficients $\boldsymbol{c}^o$ obtained in four vertical wells.

8.2 Sampling from the posterior model

The synthetic data from Section 8.1 are used as observations of a faulted reservoir. Figure 25 shows an E-W oriented vertical cross section through the seismic data at position $y = 40$. Based on the observations $\boldsymbol{o} = (\boldsymbol{s}^o, \boldsymbol{c}^o)$, four realizations of the fault pattern in the reservoir are generated from the posterior pdf $f(h|\boldsymbol{o})$, Expression (34). Each realization is generated by restarting the McMC algorithm, thus all samples are independent. 50 000 iterations are run, using an initial burn-in of 5000 for the sampling from $f(\boldsymbol{e})$. In Figure 26 the generated horizons are compared with the true, known faulted horizon. Only the part of the horizons where observations are available is included in the plots. Figure 27 gives a comparison of the true and sampled fault trace pattern on the horizon. One of the main features of the faulted horizon is a N-S striking fault with a relatively large offset in the upper part of the image, see Figure 26. This fault is present in all four realizations. A NW-SE striking fault gives a low altitude of the horizon in the lower left corner of the image. This low-altitude area is also recognized in the realizations, but there is some uncertainty about the orientation of the fault. The lower part of the N-S striking fault and the intersecting fault segment to the right have small offsets, and the realizations of edge images in Figure 27 reveal uncertainties in these areas. However, from Figure 26 the variation of altitude in these areas is observed to be relatively small both for the true faulted horizon and the realizations from the posterior.

In Figure 28 the seismic observations $\boldsymbol{s}^o$ from Figure 25 are transformed back to their

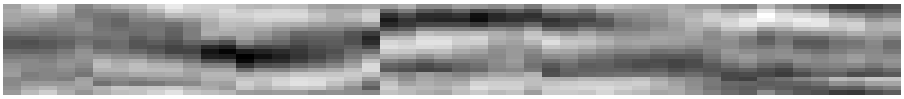
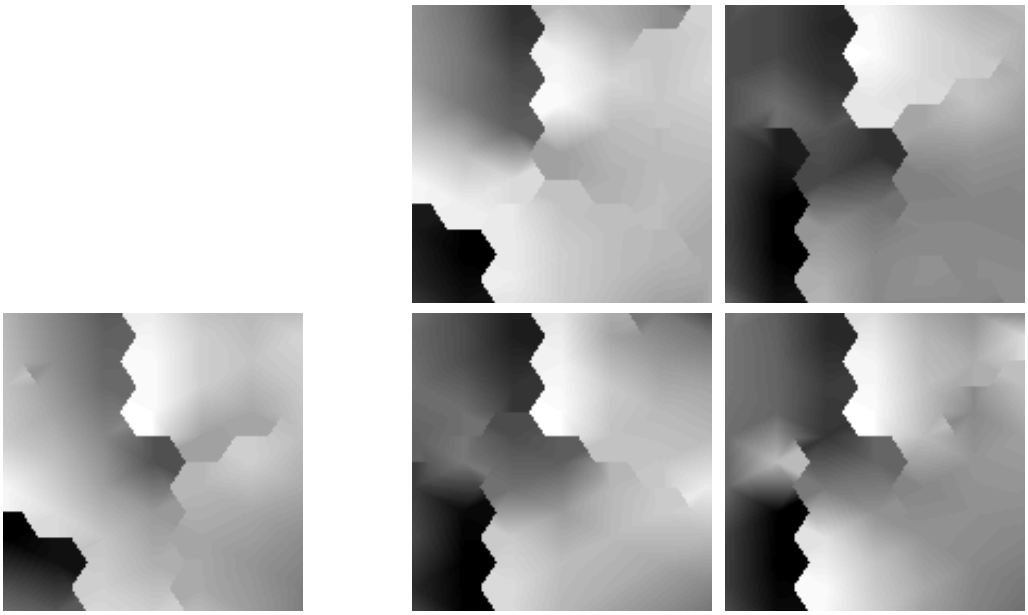
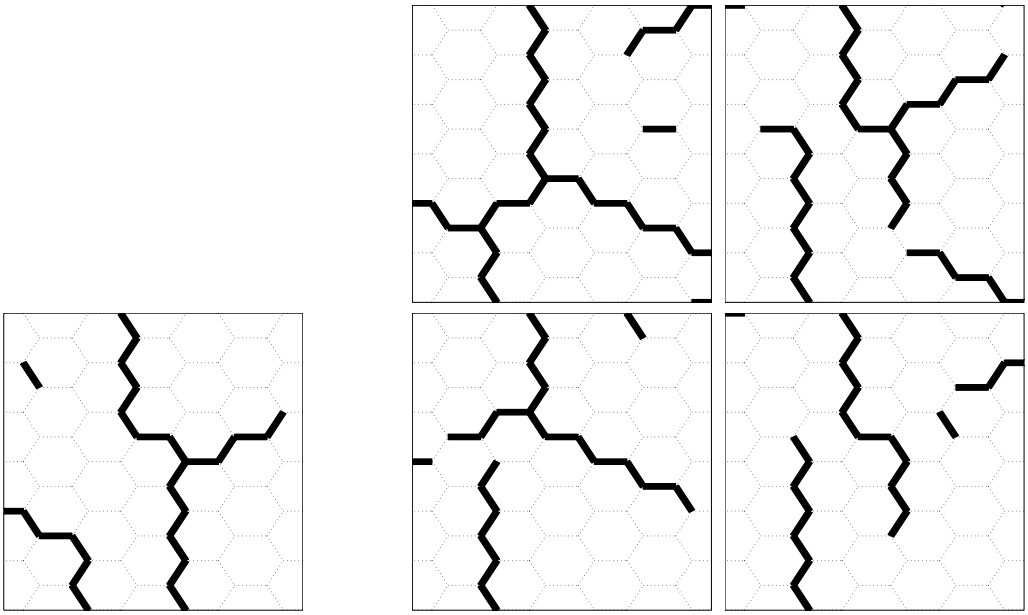


Figure 25: E-W oriented cross section through the seismic observations $\boldsymbol{s}^o$ at position $y = 40$.



True faulted horizon Realizations of the faulted horizon

Figure 26: True faulted horizon h compared to four realizations from the posterior pdf $f(h|\boldsymbol{o})$ in Expression (34).



True edge image Realizations of the edge image

Figure 27: True edge image $\boldsymbol{e}$, representing the pattern of fault traces on the horizon h, compared to the edge images of the realizations shown in Figure 26.

original structure according to the realizations of the faulted horizon in Figure 26, as illustrated in Figure 16. The cross-sections intersect the upper part of the N-S striking fault. This fault can be observed in Figure 25, while it is no longer visible in the transformed observations in Figure 28.

Figure 29a illustrates the variation in sampled offsets along fault traces. For each fault trace falling inside the observation area, the offset is measured at vertices and edge center points along the trace. The offsets of the true faulted horizon are compared to the offsets of the realizations, using different symbols for each realization. Absolute values of the offsets are reported. Fault traces that are present only in the true fault pattern or in a realization appears as points along the axes. The mismatches in Figure 29a mainly originate from the uncertainty around the position and orientation of the NW-SE striking fault in the bottom left corner of the true faulted horizon in Figure 26. This fault has a relatively large offset. Apart from offsets along these fault traces, the bulk of the points in Figure 29a are scattered around the diagonal, indicating a relatively good agreement between offsets along fault traces in the true faulted horizon and the realizations. Figure 29b shows a histogram of offset values in cases where the edge value differs between the true image $\boldsymbol{e}$ and a realization $\boldsymbol{e}_j$, see Figure 27. The figure shows that the amount of misclassified edges decreases as the offset increases, and edges that differ between the true image and a realization have mainly small offsets.

The convergence of the McMC sampling is studied by considering the energy of the likelihood function:

$$U(\boldsymbol{o}|h) = (\boldsymbol{o} - \mu_o(h))'\Sigma_o(h)^{-1}(\boldsymbol{o} - \mu_o(h))/2, \tag{37}$$

where a relatively low energy gives a relatively high likelihood, and vice versa. Figure 30 shows a plot of $U(\boldsymbol{o}|h)$ for the four runs, with accepted proposals of H indicated. It is observed from the figure that the acceptance ratio is very low. This can partly be related to the proposal distribution, which is chosen as the prior pdf. The prior pdf of this high dimensional variable will typically be much more spread out than the posterior pdf, and the proportion of proposals located in the high density area of the posterior is low. For all runs of the algorithm, the plot shows an initial decrease in the energy $U(\boldsymbol{o}|h)$ during the first 20 000 iterations. Later, few new proposals are accepted, and the energy remains low.

Each realization reported above required 1–2 days of simulation, where the McMC algorithm is implemented in the C programming language and was run on a Solaris 2 UNIX work station. As mentioned above, the time consumption of the algorithm is influenced by the time required to span the sample space of H and the time it takes to evaluate the likelihood function at each step of the algorithm. The potential for a speed up of the algorithm lies mainly within the choice of updates in the McMC algorithm and the evaluation of the approximated likelihood function at each step of the algorithm.

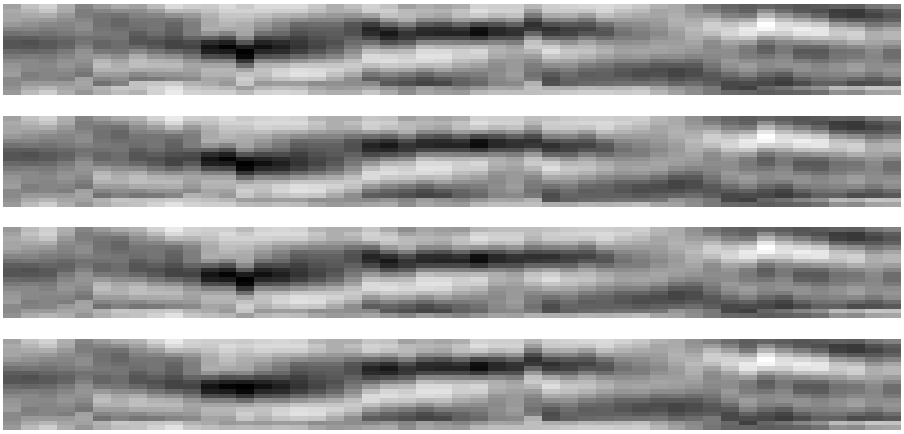


Figure 28: Seismic observations $\boldsymbol{s}^o$ from Figure 25 transformed back to the original structure, according to the realizations of the faulted horizon in Figure 26. The concept of this transformation is illustrated in Figure 16.

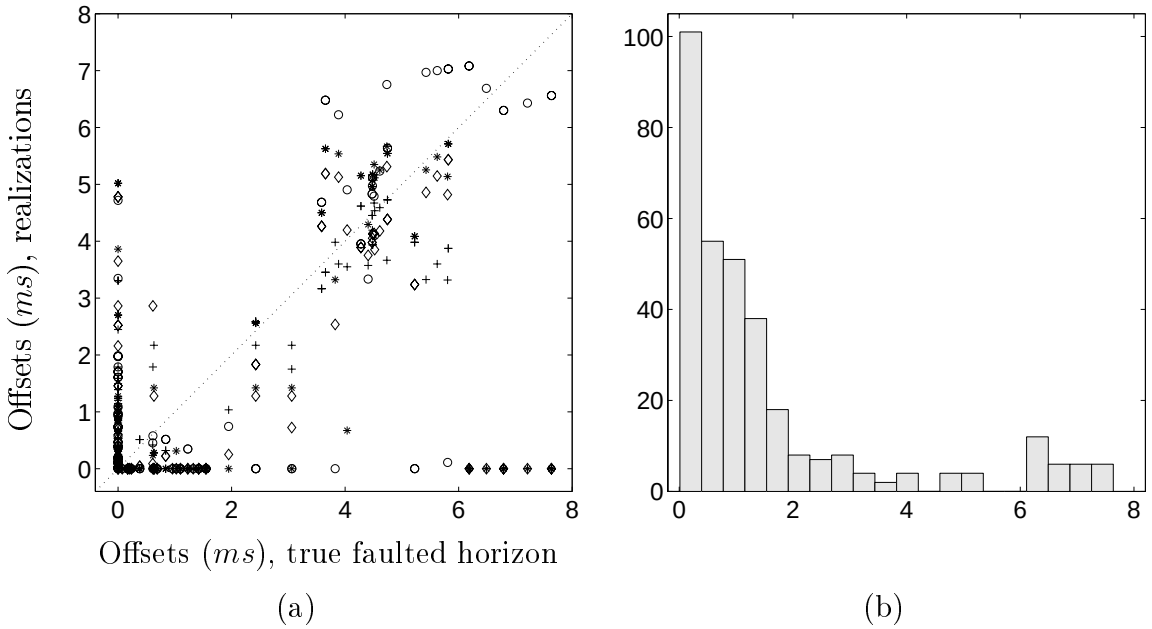


Figure 29: (a) Comparison of offset at vertices and edges for the true faulted horizon and the realizations in Figure 26. Different symbols are used for the four realizations. (b) Histogram of offset values at edges that are misclassified.

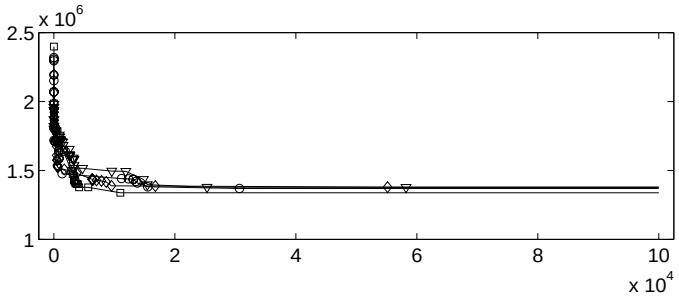


Figure 30: Plots of the energy function $U(\boldsymbol{o}|h)$ in the likelihood function $f(\boldsymbol{o}|h)$, see Expression (37), for the simulations from the posterior pdf. Different symbols ($\square$, $\diamond$, $\circ$, $\bigtriangledown$) are used to mark accepted proposals in each run of the McMC algorithm.

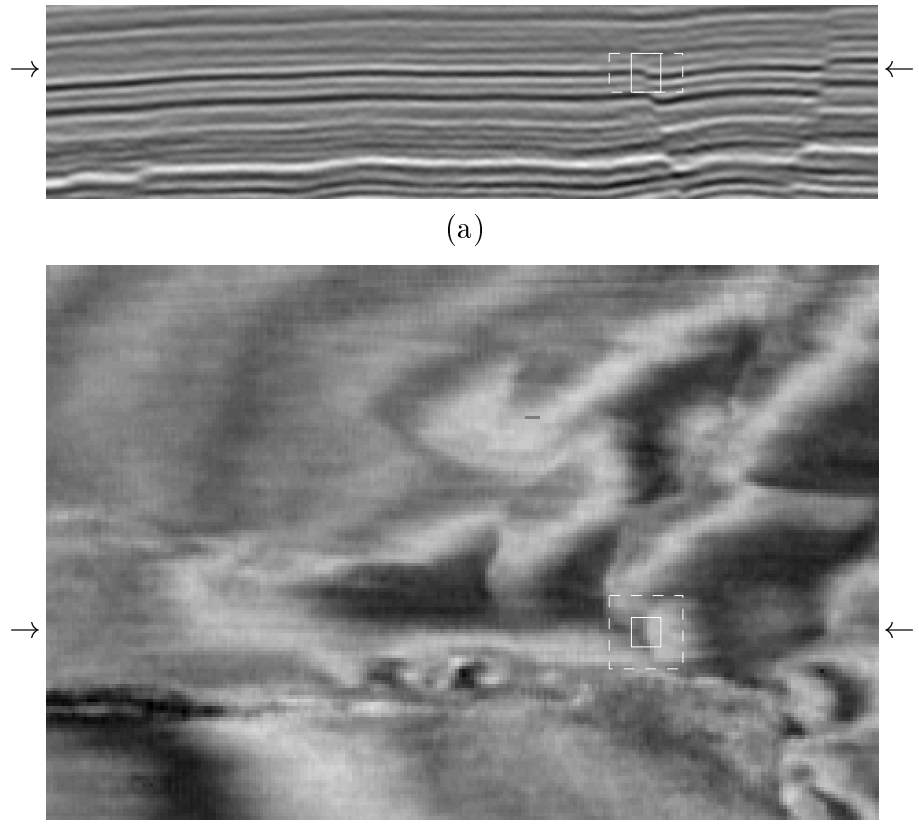
9 Real Data

In this section the stochastic model for fault patterns and the McMC sampling algorithm are applied to a dataset of real seismic observations. Only seismic data are included, since no well observations are available. The location of the seismic survey is confidential. An E-W oriented vertical cross section through the three dimensional seismic cube is presented in Figure 31a, where faults of approximately vertical dip directions are visible. Figure 31b shows a horizontal cross section through the seismic cube. The dataset is considered by geologists to be of good quality.

When the model and the McMC algorithm were first applied to large sections of this dataset, problems were encountered and no satisfactory results were obtained. To be able to investigate the nature of the problems and the limitations of the model and the algorithm, smaller sections of the dataset are studied.

A section of size $8 \times 8 \times 20$ is extracted from the data, see Figure 32, and is used in the stochastic model. The location of the extracted data is indicated in Figure 31, and the dataset is observed to contain at least one fault. Figure 33 shows realizations of the horizon H and the corresponding pattern of fault traces, $\boldsymbol{E}$, sampled from the posterior pdf in Expression (34). The number of concentric bands of pixels in the hexagonal image is 5. The parameter μ_v of the prior pdf $f(\boldsymbol{v}|\boldsymbol{e})$ is chosen as $\mu_v = 0.0025$, the potentials γ_j in the prior pdf $f(\boldsymbol{e})$ are chosen equal to the potentials of example 3 in Table 3, and the remaining variables are chosen as reported in Table 5. The number of iterations is 100 000, using 5000 initial burn-in iterations in the McMC sampling from $f(\boldsymbol{e})$.

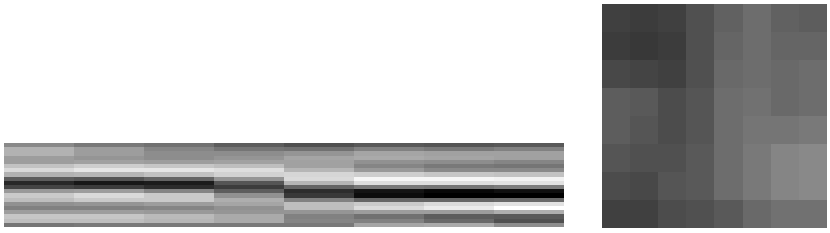
All realizations in Figure 33 show a N-S striking fault, with rock on the east side of the fault offset downwards. The realizations reveal some uncertainty in the northern part of the major fault trace and in the pattern of smaller fault traces east and west of the major fault. Figure 34 shows the original structure of the cross section in Figure 32a, obtained based



(a)

(b)

Figure 31: (a) E-W oriented vertical cross section and (b) horizontal cross section through a three dimensional seismic cube of size $151 \times 226 \times 100$. The arrows mark the positions of the cross sections. The white rectangles mark smaller sections which are extracted and further examined, one of size $8 \times 8 \times 20$ (solid lines) and one of size $20 \times 20 \times 20$ (dashed lines).



(a) (b)

Figure 32: (a) E-W oriented vertical cross section and (b) horizontal cross section through a three dimensional seismic cube of size $8 \times 8 \times 20$. The location of the dataset is marked in Figure 31.

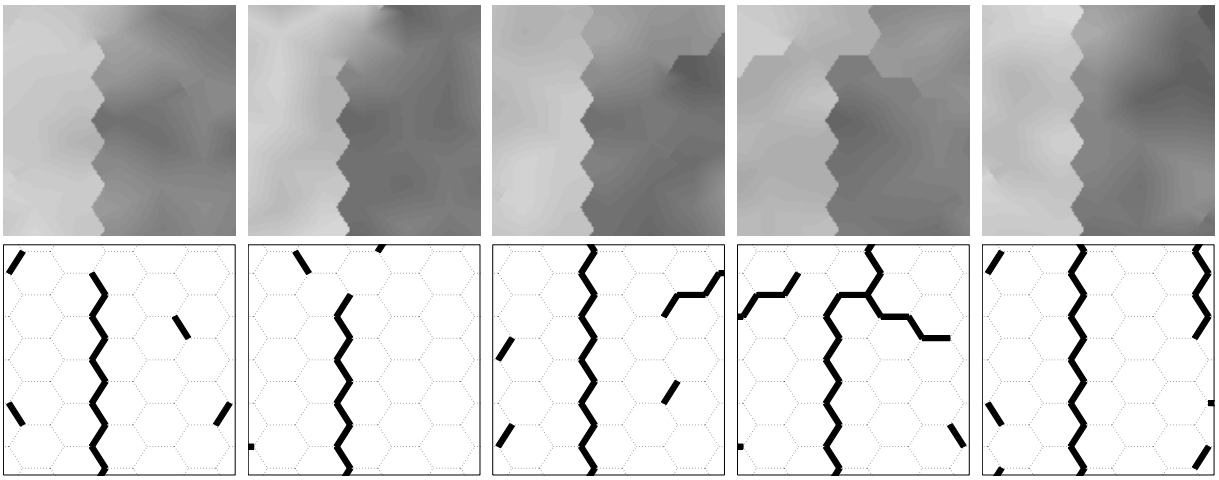


Figure 33: Realizations of H and corresponding $\boldsymbol{E}$, conditioned to a seismic cube of size $8 \times 8 \times 20$.

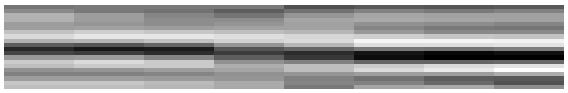


Figure 34: The vertical cross section from Figure 32a after transformation back to the original structure according to one of the realizations in Figure 33.

on one of the realizations of H in Figure 33. The vertically dipping fault in Figure 32a is no longer detectable in Figure 34. The other realizations of H give similar results. Based on the results in Figures 33 and 34, the model and the sampling algorithm are considered to work properly when applied to a small section of seismic data.

The extracted dataset is enlarged horizontally, obtaining a dataset of size $20 \times 20 \times 20$, while the hexagonal image is expanded to 11 bands of pixels. The location of the dataset is marked in Figure 31. Figure 35 shows two realizations of H and the corresponding $\boldsymbol{E}$, using the same model parameters as above and running 50 000 and 500 000 iterations of the McMC algorithm. The time consumption of the latter run was 2-3 days. No common fault traces are observed in the two realizations in Figure 35, and based on the knowledge about the true fault pattern the results of neither run seem reliable. Repeated runs with 50 000 iterations gave similar results, with no common major features in the resulting fault patterns. In general it can be concluded for this dataset that either the model assumptions are not able to reflect the reality, or that 500 000 iterations is not sufficient to give a satisfactory convergence of the McMC algorithm.

Other datasets of sizes ranging from $20 \times 20 \times 20$ to $50 \times 50 \times 20$ were extracted from the seismic cube and used in the stochastic model, and also these gave disappointing results. Most of the McMC runs resulted in horizontal, non-faulted realizations of H, although the

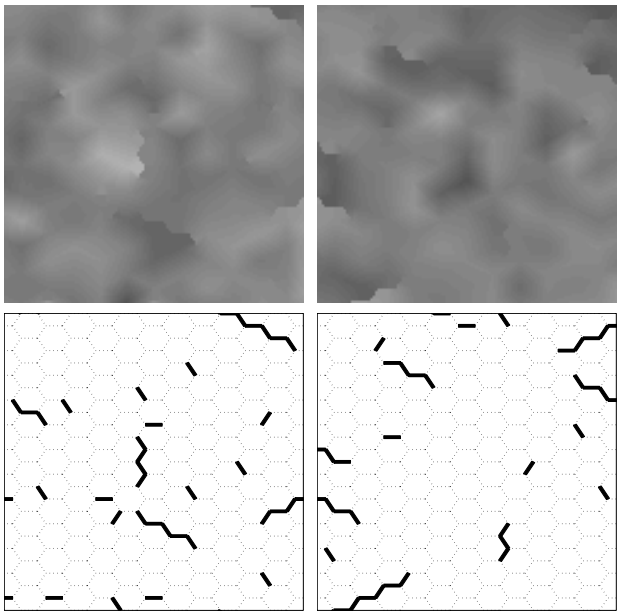


Figure 35: Realizations of H and the corresponding $\boldsymbol{E}$, conditioned to a seismic cube of size $20 \times 20 \times 20$.

datasets were extracted from regions where faults are visible in the seismic data. Relatively coarse grids of 5-7 bands of pixels were used for the hexagonal image.

The examples above show satisfactory results when only a small seismic cube is used in the model, while problems arise when the dataset is enlarged. The main sources of the problems encountered are believed to be restrictive model assumptions and the time consumption of the sampling algorithm. The time consumption of each iteration of the McMC algorithm depends on the grid sizes of both the hexagonal image and the seismic cube, which determines the time required to sample from the prior and to evaluate the likelihood. The model assumptions that are considered to be potential sources of difficulties are:

(i) Faults are vertical with constant offset vertically.
(ii) Fault planes are infinitely thin.
(iii) Offsets are small compared to the thickness of the formation.
(iv) All fault traces follow the edges of the hexagonal grid.
(v) Reflection coefficients and seismic data follow Gaussian distributions.

The mean and variance of the seismic data of a faulted reservoir are obtained under the assumption of vertical fault planes. If this model is applied to datasets with non-vertical fault planes, the mean and variance will not be obtained correctly as the transformation of rock due to faulting is not perfectly modeled. It is observed in Figure 31a that the faults are only approximately vertical and the offset seems to vary somewhat vertically,

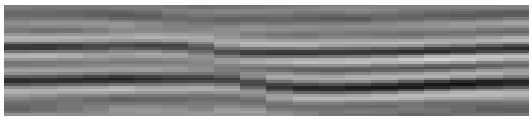


Figure 36: E-W oriented vertical cross section of size 20 (horizontally) $\times$ 40 extracted from the cross section shown in Figure 31a. The cross section covers an enlarged area around the data in Figure 32a.

thus assumption (i) is not completely fulfilled. However, within the extracted dataset in Figure 32a the fault plane is vertical and the offset is constant vertically. Figure 36 shows 20×40 seismic datapoints, where the cross section in Figure 32a is enlarged both horizontally and vertically. From this cross section it is observed that the fault plane does not appear as a vertical, well defined discontinuity, but rather as a dipping band of deformations covering 5-6 observation points horizontally. Thus the figure indicates violations also of assumption (ii). Furthermore, the seismic data close to the fault seem to be somewhat blurred, which may cause slight violations of assumption (v). The blurring is related to the Fresnel zones of the seismic signal, see Sheriff and Geldart (1995). The seismic signal is in practice not reflected from a single point on a reflector, but from a larger area. Close to a fault plane the Fresnel zone will cover areas on both sides of the fault, and the seismic signal from a specific reflector is weakened.

Seismic datasets consisting of 20 datapoints vertically have been used in the examples reported above, corresponding to a thickness of the formation of $40ms$. This vertical thickness enables modeling of faults with vertical offsets up to approximately $20ms$. In the real dataset studied, offset values range from about $4ms$ to beyond $60ms$, thus with seismic cubes of thickness $40ms$ assumption (iii) is not fulfilled by all faults. An example of large offset faults is given in Figure 37. To be able to model the largest fault offsets, seismic cubes of vertical thickness $120 - 150ms$, or 60-75 datapoints, should be used. This magnifies the seismic cube by a factor 3-4 compared to the previous examples, increasing the time required to evaluate the likelihood. Furthermore, the deviations from vertical fault planes become more pronounced when thicker seismic cubes are used.

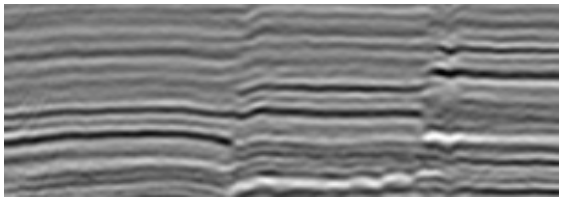


Figure 37: N-S oriented vertical cross section through the three dimensional seismic dataset presented in Figure 31. The two largest faults have offsets of $30 - 50ms$.

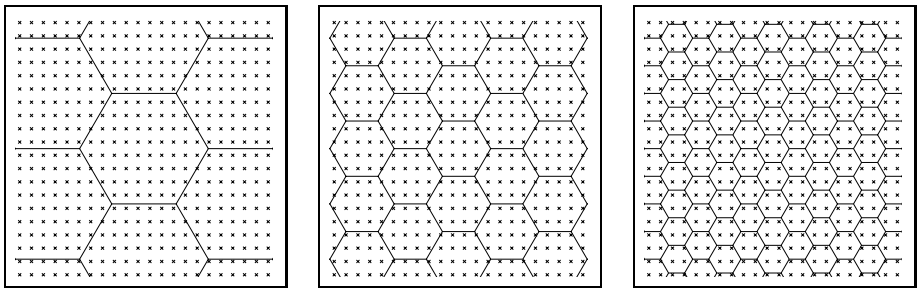


Figure 38: Seismic datapoints (crosses), combined with hexagonal images of gradually finer grids.

Assumption (iv) is obviously not fulfilled by the real dataset. The flexibility in modeling the positions of fault traces relative to the seismic observations depends on the resolution of the hexagonal grid. This is illustrated in Figure 38. The coarse grid in Figure 38a results in large blocks of seismic data inside each hexagonal pixel, with zero probability of any fault traces dividing these seismic datapoints. Figures 38b and 38c show how finer grids can be used to improve the flexibility in the modeling of fault traces. However, refinement of the hexagonal grid implies an enlargement of the sample space of $(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$, increasing the number of McMC iterations required to span the sample space properly.

A small test is performed aiming at exposing problems related to assumption (v), by investigating the performance of the model and algorithm when assumptions (i)–(iv) are fulfilled. A seismic cube of size $30 \times 30 \times 20$ with no visible faults is used. Two simulation cases are performed, one with the original data and one with a fault pattern sampled from the prior model artificially enforced on the dataset. A hexagonal grid with 5 concentric bands of pixels is used. Simulations based on the original data gave approximately horizontal realizations of H with no faults at seismic scales. When a N-S striking fault was artificially enforced, see Figure 39a, all realizations from the posterior pdf reproduced the fault, see Figure 39b. Based on these examples, no major problems with the Gaussian assumption are apparent.

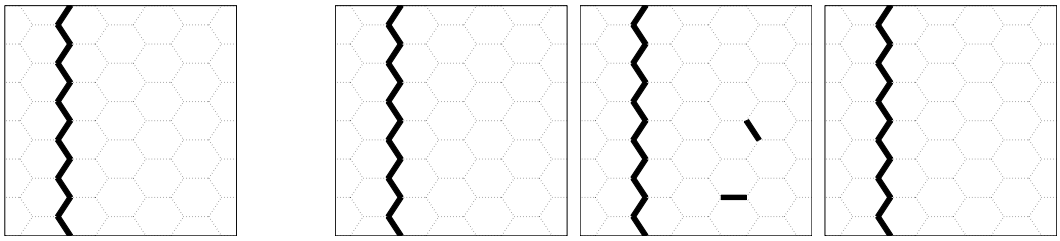


Figure 39: (a) Fault trace artificially enforced on a seismic cube. (b) Realizations from the posterior pdf.

Based on the experience with the real data, it is clear that with the computer power currently available, the performance of the model and the McMC algorithm is of limited value for large real datasets. Model assumptions (iii) and (iv) are related to the time consumption of the McMC algorithm, and problems related to these assumptions could be improved on by increased computer power or more efficient implementation, for example by the use of parallel implementation to evaluate the likelihood function. Violations of assumptions (i) and (ii) can be corrected for by redefining the model to increase flexibility in the modeling of the faults, while assumption (v) is not found to cause any major problems. A possible application for large datasets could be to apply the model to a smaller section within a gliding window. By moving the window across the larger dataset and modeling the fault pattern within a number of smaller sections, the overall fault pattern could be modeled. This could be done either by modeling the fault pattern within each section independently or by including dependencies between the sections, for example modeled through boundary conditions. An alternative approach to dealing with large datasets is to apply a thinning of the seismic data in the horizontal direction. This could improve the fulfillment of assumptions (i) and (ii), and would reduce the size of the seismic cube and thus the time consumption of each McMC iteration. Furthermore, the thinning would have a similar effect on assumption (iv) as the refinement of the grid illustrated in Figure 38. However, the thinning will reduce the seismic resolution horizontally, and can cause problems in distinguishing between fault discontinuities and continuous changes in elevation of comparable magnitude.

10 Conclusion

The paper presents a stochastic model for patterns of faults above seismic resolution, conditioned to three dimensional seismic data and well observations. Based on the typically strong horizontal correlation found in reflection coefficients, and also in seismic data arising from a linear convolution of the reflection coefficients, faults can be recognized as discontinuities in the data.

Due to the complexity of the model, McMC techniques are required to sample from the posterior distribution of fault patterns. The performance of the model and the constructed McMC algorithm is illustrated using a synthetic dataset. In this case the samples from the posterior distribution successfully reproduce the main features of the underlying, known fault pattern.

Problems are encountered when the model and the sampling algorithm are applied to a real seismic dataset, and possible sources of difficulties are discussed. Some of the model assumptions are found not to be completely satisfied. The flexibility of the model is influenced by the resolution of the model, however refinement of the grid is obtained on the expense of the time consumption. Redefinition of the model, increased computer

power and more efficient implementation are suggested as improvements of the problems with the model assumptions and resolution. Alternatively, the model and the algorithm can be applied to smaller sections of the data, for example by the use of a gliding window or by thinning of the seismic data.

References

Abrahamsen, P. (1997). A Review of Gaussian Random Fields and Correlation Functions, *Report 917*, Norwegian Computing Center. Second edition.

Badley, M. E., Freeman, B., Roberts, A. M., Thatcher, J. S., Walsh, J., Watterson, J. and Yielding, G. (1990). Fault Interpretation During Seismic Interpretation and Reservoir Evaluation, *The Integration of Geology, Geophysics, Petrophysics and Petroleum Engineering in Reservoir Delineation, Description and Management*, Proceedings of the 1st Archie Conference, Houston, Texas, pp. 224–241. AAPG.

Bahorich, M. S. and Farmer, S. L. (1995). The coherence cube, *The Leading Edge* **14**(10): 1053–1058.

Bahorich, M. S. and Farmer, S. L. (1996). Method of seismic signal processing and exploration. U. S. Patent Number 5,563,949.

Belfield, W. C. (1998). Incorporating spatial distribution into stochastic modelling of fractures: multifractal and levý-stable statistics, *Journal of Structural Geology* **20**(4): 473–486.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems, *J. R. Statist. Soc. B* **36**(2): 192–236.

Besag, J. (1989). Towards Bayesian image analysis, *Journal of Applied Probability* **16**(3): 395–407.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion), *Statistical Science* **10**(1): 3–66.

Brand, P. J. and Haldorsen, H. H. (1988). Modelling large-scale heterogeneities caused by faulting with a stochastic approach, *Revue de l'Institut français du pétrole* **43**(5): 647–657.

Chilès, J. P. (1988). Fractal and Geostatistical Methods for Modeling of a Fracture Network, *Mathematical Geology* **20**(6): 631–654.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, Wiley, New York.

Dawers, N. H., Anders, M. H. and Scholz, C. H. (1993). Growth of normal faults: Displacement-length scaling, *Geology* **21**: 1107–1110.

Eide, A. L. (1999). *Stochastic Reservoir Characterization Constrained by Seismic Data*, PhD thesis, Norwegian University of Science and Technology. Doktoringeniøravhandling 1999:5.

Eide, A. L., Omre, H. and Ursin, B. (1997a). Stochastic reservoir characterization conditioned on seismic data, *in* E. Y. Baafi and N. A. Schofield (eds), *Geostatistics Wollongong '96*, Vol. 1, Kluwer Academic Publishers, pp. 442–453.

Eide, A. L., Ursin, B. and Omre, H. (1997b). Stochastic simulation of porosity and acoustic impedance conditioned to seismic data and well data, *67th Annual Meeting of the Society of Exploration Geophysicists, Expanded Abstracts*, SEG, Dallas, Texas.

Gauthier, B. D. M. and Lake, S. D. (1993). Probabilistic Modeling of Faults Below the Limit of Seismic Resolution in Pelican Field, North Sea, Offshore United Kingdom, *The American Association of Petroleum Geologists Bulletin* **77**(5): 761–777.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6): 721–741.

Gringarten, E. (1996). 3-D Geometric Description of Fractured Reservoirs, *Mathematical Geology* **28**(7): 881–893.

Gringarten, E. (1998). Fracnet: stochastic simulation of fractures in layered systems, *Computers and Geosciences* **24**(8): 729–736.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrica* **57**: 97–109.

Lia, O., Omre, H., Tjelmeland, H., Holden, L. and Egeland, T. (1997). Uncertainties in Reservoir Production Forecasts, *AAPG Bulletin* **81**(5): 775–802.

Luo, Y., Higgs, W. G. and Kowalik, W. S. (1996). Edge detection and stratigraphic analysis using 3d seismic data, *66th Annual Meeting of the Society of Exploration Geophysicists, Expanded Abstracts*, SEG, Denver, Colorado, pp. 324–327.

Marfurt, K. J., Kirlin, R. L., Farmer, S. L. and Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm, *Geophysics* **63**(4): 1150–1165.

Marfurt, K. J., Sudhaker, V., Gersztenkorn, A., Crawford, K. D. and Nissen, S. E. (1999). Coherency calculations in the presence of structural dip, *Geophysics* **64**(1): 104–111.

Monroe, J. S. and Wicander, R. (1994). *The Changing Earth*, West Publishing Company.

Munthe, K., Holden, L., Mostad, P. and Townsend, C. (1994). Modelling Sub-seismic Fault Patterns Using a Marked Point Process, *Proceedings of the 4th European conference on the Mathematics of Oil Recovery, Røros, Norway.* Topic B: Heterogeneity Description and Assessment of Uncertainty.

Munthe, K. L., Omre, H., Holden, L., Damsleth, E., Heffer, K., Olsen, T. S. and Watterson, J. (1993). Subseismic Faults in Reservoir Description and Simulation, *68th Annual Conference and Exhibition of the Society of Petroleum Engineers*, Society of Petroleum Engineers, pp. 843–850. SPE Paper 26500.

Omre, H. and Tjelmeland, H. (1997). Petroleum geostatistics, *in* E. Y. Baafi and N. A. Schofield (eds), *Geostatistics Wollongong '96*, Vol. 1, Kluwer Academic Publishers, pp. 41–52.

Ripley, B. D. (1987). *Stochastic Simulation*, Wiley, New York.

Sheriff, R. E. and Geldart, L. P. (1995). *Exploration Seismology*, 2 edn, Cambridge University Press.

Tjelmeland, H. (1996). *Stochastic models in reservoir characterization and Markov random fields for compact objects*, PhD thesis, Norwegian University of Science and Technology. Doktoringeniøravhandling 1996:44.

Tjelmeland, H. and Besag, J. (1998). Markov Random Fields with Higher-Order Interactions, *Scandinavian Journal of Statistics* **25**(3): 415–433.

Todoeschuck, J. P., Jensen, O. G. and Labonte, S. (1990). Gaussian scaling noise model of seismic reflection sequences: Evidence from well logs, *Geophysics* **55**(4): 480–484.

Walsh, J. J. and Watterson, J. (1987). Distribution of cumulative displacement and seismic slip on a single normal fault surface, *Journal of Structural Geology* **9**(8): 1039–1046.

Walsh, J. J. and Watterson, J. (1988). Analysis of the relationship between displacements and dimensions of faults, *Journal of Structural Geology* **10**(3): 239–247.

Wen, R. and Sinding-Larsen, R. (1997). Stochastic modeling and simulation of small faults by marked point processes and kriging, *in* E. Y. Baafi and N. A. Schofield (eds), *Geostatistics Wollongong '96*, Vol. 1, Kluwer Academic Publishers, pp. 398–414.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer.

A Sampling Algorithms

A Metropolis-Hastings algorithm is used to sample from the posterior pdf $f(h|\boldsymbol{o})$ of the faulted horizon H conditioned to the observations $\boldsymbol{O}$ of the reservoir. Several alternative transition steps are used, resulting in different transition kernels. The algorithm is given as follows:

Algorithm 1 — Sampling from $f(h|\boldsymbol{o})$

Let n_e, n_v and n_p be the number of edges $\boldsymbol{E}$, vertices $\boldsymbol{V}$ and pixels $\boldsymbol{P}$ respectively, and let $\Omega_e^{n_e}$, $\Omega_v^{n_v}$ and $\Omega_p^{n_p}$ be the sample spaces of $\boldsymbol{E}$, $\boldsymbol{V}$ and $\boldsymbol{P}$. Furthermore, let $\sum_{i=1}^{3} \pi_i = 1$, where $0 \leq \pi_i \leq 1$.

- *Initiate arbitrary $\boldsymbol{e}(0) \in \Omega_e^{n_e}$, $\boldsymbol{v}(0) \in \Omega_v^{n_v}$ and $\boldsymbol{p}(0) \in \Omega_p^{n_p}$, let $h(0) = h(\boldsymbol{e}(0), \boldsymbol{v}(0), \boldsymbol{p}(0))$.*
- *Iterate $t = 1, 2, \ldots$*
  - *Let $\boldsymbol{e} = \boldsymbol{e}(t-1)$, $\boldsymbol{v} = \boldsymbol{v}(t-1)$, $\boldsymbol{p} = \boldsymbol{p}(t-1)$ and $h = h(t-1)$.*
  - *Generate $\boldsymbol{e}'$, $\boldsymbol{v}'$ and $\boldsymbol{p}'$ in one of the following ways:*

    (i) *With probability π_1:*
    $$\begin{aligned} & \boldsymbol{e}' \sim f(\boldsymbol{e}') \\ & \boldsymbol{v}' \sim f(\boldsymbol{v}'|\boldsymbol{e}') \\ & \boldsymbol{p}' \sim f(\boldsymbol{p}'|\boldsymbol{e}', \boldsymbol{v}') \\ \textit{hence} \quad & q(h'|h) = f(\boldsymbol{e}')f(\boldsymbol{v}'|\boldsymbol{e}')f(\boldsymbol{p}'|\boldsymbol{e}', \boldsymbol{v}') \end{aligned}$$

    (ii) *With probability π_2:*
    $$\begin{aligned} & \boldsymbol{e}' = \boldsymbol{e} \\ & \boldsymbol{v}' \sim f(\boldsymbol{v}'|\boldsymbol{e}) \\ & \boldsymbol{p}' \sim f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}') \\ \textit{hence} \quad & q(h'|h) = f(\boldsymbol{v}'|\boldsymbol{e})f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}') \end{aligned}$$

    (iii) *With probability π_3:*
    $$\begin{aligned} & \boldsymbol{e}' = \boldsymbol{e} \\ & \boldsymbol{v}' = \boldsymbol{v} \\ & \boldsymbol{p}' \sim f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}) \\ \textit{hence} \quad & q(h'|h) = f(\boldsymbol{p}'|\boldsymbol{e}, \boldsymbol{v}) \end{aligned}$$

  - *$h' = h'(\boldsymbol{e}', \boldsymbol{v}', \boldsymbol{p}')$.*
  - *With acceptance probability $\alpha(h'|h)$*

    $$\alpha(h'|h) = \min\left\{1, \ \frac{f(\boldsymbol{o}|h')}{f(\boldsymbol{o}|h)}\right\}$$

    let $h(t) = h'$, $\boldsymbol{e}(t) = \boldsymbol{e}'$, $\boldsymbol{v}(t) = \boldsymbol{v}'$ and $\boldsymbol{p}(t) = \boldsymbol{p}'$.
    Otherwise let $h(t) = h$, $\boldsymbol{e}(t) = \boldsymbol{e}$, $\boldsymbol{v}(t) = \boldsymbol{v}$ and $\boldsymbol{p}(t) = \boldsymbol{p}$.

The variable $H(t)$ has pdf $f^t(h|\boldsymbol{o})$ where $\lim_{t\to\infty} f^t(h|\boldsymbol{o}) = f(h|\boldsymbol{o})$.

Algorithm 1 includes steps where realizations are generated from the prior pdfs $f(\boldsymbol{e})$, $f(\boldsymbol{v}|\boldsymbol{e})$ and $f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v})$. Exact samples can be drawn from the prior distributions of $\boldsymbol{V}$ and $\boldsymbol{P}$, while McMC techniques are used to generate samples from the prior distribution of $\boldsymbol{E}$. Algorithms 2–4 below describe how samples can be generated from the prior distributions.

The edges $\boldsymbol{E}$ are modeled as a Markov random field, see Section 3.2.1. A sample from the prior pdf $f(\boldsymbol{e})$ of the edges is generated using the following Metropolis-Hastings algorithm:

Algorithm 2 — Sampling from $f(\boldsymbol{e})$

Let n_e be the number of edges in $\boldsymbol{E}$ and $\Omega_e^{n_e}$ the sample space of $\boldsymbol{E}$.

- *Initiate arbitrary* $\boldsymbol{e}(0) \in \Omega_e^{n_e}$.
- *Iterate* $t = 1, 2, \ldots$
  - $\boldsymbol{e} = \boldsymbol{e}(t-1)$.
  - *For* $k = 1, 2, \ldots, n_e$
    - $i \sim \text{Unif}\{1, \ldots, n_e\}$.
    - $e_i' \sim \text{Unif}\{\Omega_e \setminus \{e_i\}\}$, $e_j' = e_j$ *for* $j \neq i$.
    - *Hence* $q(\boldsymbol{e}'|\boldsymbol{e}) = n_e^{-1}(|\Omega_e| - 1)^{-1}$.
    - *With acceptance probability*

$$\alpha(\boldsymbol{e}'|\boldsymbol{e}) = \min\left\{1,\ \exp\{-\sum_{c \in \mathcal{C}; i \in c} (\omega_c(\boldsymbol{e}_c') - \omega_c(\boldsymbol{e}_c))\}\right\},$$

      let $e_i = e_i'$, *otherwise retain* e_i.
  - $\boldsymbol{e}(t) = \boldsymbol{e}$.

The edge values $\boldsymbol{E}(t)$ has pdf $f^t(\boldsymbol{e})$ where $\lim_{t\to\infty} f^t(\boldsymbol{e}) = f(\boldsymbol{e})$.

When Algorithm 2 is used within Algorithm 1, burn-in iterations of Algorithm 2 is run before Algorithm 1 is started. When a realization from $f(\boldsymbol{e})$ is needed in update step (i) of Algorithm 1, a limited number of iterations of Algorithm 2 is performed. These iterations of Algorithm 2 are performed to reduce dependency between the samples from $f(\boldsymbol{e})$. Typically, 5–10 000 burn-in iterations of Algorithm 2 are run, while realizations from $f(\boldsymbol{e})$ used in Algorithm 1 are separated by 50 iterations of Algorithm 2.

The prior distribution of the vertices $\boldsymbol{V}$, conditioned on the edges $\boldsymbol{E}$, is a combination of a stochastic offset direction along each fault trace and a deterministic offset profile given the offset direction, see Section 3.2.2. A sample from the prior pdf $f(\boldsymbol{v}|\boldsymbol{e})$ is generated by exact drawing of the offset directions and evaluation of the corresponding offset profiles. The algorithm is given as follows:

Algorithm 3 — Sampling from $f(\boldsymbol{v}|\boldsymbol{e})$

Suppose $\sum_{b_j \in \mathcal{B}_e} |b_j| + |b_0| = n_v$, *where* n_v *is the number of vertices in* $\boldsymbol{V}$, b_j *is the index set of vertices along fault trace* j *and* $\mathcal{B}_e$ *is a partitioning of edges in the realization* $\boldsymbol{e}$ *into fault traces. The sample space of* $\boldsymbol{V}$ *is given as* $\Omega_v^{n_v} \subseteq \mathbb{R}^{n_v}$.

- *For* $i \in b_0$
  - $V_i = 0$
- *For* $b \in \mathcal{B}_e$
  - $D_b \sim f(d_b)$, *where* $f(1) = \theta$, $f(-1) = 1 - \theta$.
  - $n =$ *length of fault trace* b
  - $i =$ *number of isolated fault tips of* b
  - $\boldsymbol{V}_b = D_b \boldsymbol{g}_{i,n}$

A sample of $\boldsymbol{V}$ from the prior pdf $f(\boldsymbol{v}|\boldsymbol{e})$ is thus obtained. The vectors $\boldsymbol{g}_{i,n}$, $i = 0, 1, 2$, are defined in Section 3.2.2. If there are several possible partitionings of $\boldsymbol{e}$ into fault traces, the set $\mathcal{B}_e$ is picked at random among the $n_b(\boldsymbol{e})$ possibilities. This is done by randomly partitioning each intersection point into a fault tip and an intersected fault trace.

The pixels $\boldsymbol{P}$ are modeled as a multivariate Gaussian distribution conditioned on edges $\boldsymbol{E}$ and vertices $\boldsymbol{V}$, see Section 3.2.3. An exact sample from the prior distribution of the pixels, conditioned on the edges and vertices, is generated in the following way:

Algorithm 4 — Sampling from $f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v})$

Let n_p *be the number of pixels in* $\boldsymbol{P}$. *The prior pdf of pixels, conditioned on edges and vertices, is* $f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v}) = \phi(\boldsymbol{p}; \mu(\boldsymbol{e}, \boldsymbol{v}), \Sigma_p)$, *and the sample space of* $\boldsymbol{P}$ *is* $\Omega_p^{n_p} \subseteq \mathbb{R}^{n_p}$.

- *Find the matrix* $\boldsymbol{A}$ *using a Cholesky decomposition of* Σ_p: $\Sigma_p = \boldsymbol{A}\boldsymbol{A}'$.
- *For* $i = 1, 2, \ldots, n_p$ *generate* $Y_i \sim N(0, 1)$.
- *Let* $\boldsymbol{y} = (y_1, y_2, \ldots, y_{n_p})$ *and* $\boldsymbol{p} = \mu(\boldsymbol{e}, \boldsymbol{v}) + \boldsymbol{A}\boldsymbol{y}$.

The pixel values $\boldsymbol{p}$ form a sample from the prior pdf $f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v})$. Inversion sampling is used to sample from the standard Gaussian distribution $N(0, 1)$, see Ripley (1987).

Part V

Partitioning of a line segment

Partitioning of a line segment

Hilde G. Borgos

Abstract

Different procedures for the partitioning of a line segment is studied. An interval on the real line is partitioned according to a breakage procedure, resulting in a number of smaller segments. Conditioned on the number of resulting segments the different procedures give different length distributions of the segments. The length distributions are studied analytically.

1 Introduction

In a discussion on geological faults and fault patterns, a question turned up concerning the resulting length distribution of smaller segments, when a line segment is partitioned according to different partitioning processes. The aim of this report is to answer some of these questions, by studying some possible partitioning processes analytically.

In Section 2 the partitioning of a line segment is expressed as a partitioning of an interval on the real line. In Sections 3 through 7 different partitioning procedures are studied. The resulting distributions of the lengths of the smaller segments are found, using the notation of Section 2. In Section 8 some of the different distributions are compared graphically.

2 Partitioning of a Line Segment

A line segment of length L can be represented by an interval of length L on the real line $\mathbb{R}$. The interval is partitioned into n intervals Y_i whose lengths sum up to L. Depending on the partitioning process, the resulting intervals Y_i will follow different distributions. If $n-1$ points X_l are distributed along the line, these points, together with the endpoints of the line, define the endpoints of the intervals Y_i. Placing the line on the real line $\mathbb{R}$, with one end at the origin and the other at L, the partitioning can be expressed as follows:

- Let the line be represented by the interval $[0, L]$.
- Distribute the points $X_1, X_2, \ldots, X_{n-1}$ in the interval.
- Order the points; $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$.
- The ordered points, together with the endpoints of the line, define n intervals of lengths $Y_1, Y_2, \ldots, Y_n$, where $\sum_{i=1}^{n} Y_i = L$. These lengths are given by $Y_1 = X_{(1)}$, $Y_i = X_{(i)} - X_{(i-1)}$, $i = 2, 3, \ldots, n-1$ and $Y_n = L - X_{(n-1)}$.

3 Equidistant Partitioning Points

Let the points X_l be placed equidistantly in the interval $[0, L]$, as in Figure 1. The positions of the points are then

$$X_l = X_{(l)} = \frac{Ll}{n}, \quad l = 1, 2, \ldots, n-1,$$

where the ordered points $X_{(l)}$ are equal to the unordered points X_l. Denote the endpoints of the interval $X_{(0)} = 0$ and $X_{(n)} = L$. This yields intervals Y_i with lengths

$$Y_i = X_{(i)} - X_{(i-1)} = \frac{Li}{n} - \frac{L(i-1)}{n} = \frac{L}{n}, \quad i = 1, 2, \ldots, n.$$

All intervals have the same length, determined by the length of the line and the number of intervals, and the density function of an interval length Y is given by the Dirac's delta function,

$$f_Y(y) = \delta\left(y - \frac{L}{n}\right),$$

where $\delta(t) = 0$ for $t \neq 0$ and $\int_{-\infty}^{\infty} \delta(t)dt = 1$.

Figure 1: The $n-1$ points X_l are placed equidistantly in the interval $[0, L]$.

4 Independent Uniformly Distributed Partitioning Points

Consider a stationary Poisson point process in $\mathbb{R}$, with intensity λ (Stoyan, Kendall and Mecke, 1995). The distance between two succeeding points is then exponentially distributed

with the same parameter λ. Now, consider the finite interval $[0, L]$. Conditioned on the number of points falling in this interval, we want to find the distribution of the lengths between the points. This distribution will no longer be exponential, as all intervals have an upper bound of L. Given the number of points in the interval, the positions of the points are independent identically uniformly distributed on the interval, forming a binomial point process (Stoyan and Stoyan, 1994). Let the points $X_1, X_2, \ldots, X_{n-1}$ be iid with a uniform distribution $X_l \sim U[0, L]$. An example is shown in Figure 2. The density function of an interval length Y is then given by

$$f_Y(y) = \frac{n-1}{L} \cdot \left(1 - \frac{y}{L}\right)^{n-2}, \quad 0 \leq y \leq L. \tag{1}$$

For a proof, see Appendix A.1. The density function corresponds to the scaled length $W = Y/L$ following a beta distribution, $W \sim \text{Beta}(1, n-1)$, which is a well known result. The mean and variance of an interval length Y are

$$\text{E}[Y] = \frac{L}{n}, \qquad \text{Var}[Y] = L^2 \frac{n-1}{n^2(n+1)}. \tag{2}$$

The mean is equivalent to the lengths of the intervals when the points X_l are equidistantly distributed. If we let $L \to \infty$ and $n \to \infty$ while $n/L = \lambda$ remains constant, the limiting distribution is the exponential distribution with parameter λ. For a proof, see Appendix B.1. The limiting distribution corresponds to the X's being a stationary Poisson process on the real line.

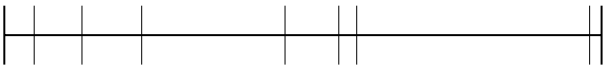


Figure 2: The $n-1$ points X_l are iid random variables with a uniform distribution $U[0, L]$.

5 Repulsion Between the Partitioning Points

Repulsion between the points X_l can be created by thinning of another point process. Let $\mathbf{Z}$ be a set of points $\mathbf{Z} = \{Z_j\}_{j=1}^{s}$, $s \geq n$, distributed on the interval $[0, L]$, and let the set of points $\mathbf{X}$ be a subset of $\mathbf{Z}$, that is $\mathbf{X} \subset \mathbf{Z}$. Let $Z_{(j)}$, $j = 1, 2, \ldots, s$, be the ordered points. If the points X_l are given by $X_l = Z_{(pl)}$, $l = 1, 2, \ldots, n-1$, where p is a positive integer, there are $p-1$ points $Z_{(pl+1)}, \ldots, Z_{(pl+p-1)}$ located between every neighbor pair of ordered points $X_{(l)}$. These $p-1$ points "force" the two points away from each other, creating repulsion. An example is shown in Figure 3. Let the endpoints of the interval be the points $X_{(0)} = 0$ and $X_{(n)} = L$. To get $p-1$ points between every $X_{(l)}$, $l = 0, 1, \ldots, n$, a total number of $s = pn - 1$ points Z_j have to be distributed along the line. If the s points Z_j are independent uniformly distributed on $[0, L]$, forming a binomial point process (Stoyan and

Stoyan, 1994), the density function of an interval length $Y_i = X_{(i)} - X_{(i-1)} = Z_{(pi)} - Z_{(p(i-1))}$ is

$$f_Y(y) = \frac{p}{L} \cdot \binom{pn-1}{p} \left(\frac{y}{L}\right)^{p-1} \left(1 - \frac{y}{L}\right)^{p(n-1)-1}, \quad 0 \le y \le L. \tag{3}$$

This is shown by Mouland (1997), and an alternative proof is given in Appendix A.2. The density function corresponds to the scaled length $W = Y/L$ following a beta distribution, $W \sim \text{Beta}(p, p(n-1))$. Substituting $p = 1$ in Expression (3) we return to Expression (1), as expected, since $p = 1$ corresponds to the X_l's being independent uniformly distributed. The mean and variance of the interval lengths in this case are

$$\text{E}[Y] = \frac{L}{n}, \qquad \text{Var}[Y] = L^2 \frac{n-1}{n^2(pn+1)}.$$

As in the previous cases, the mean equals the lengths of the intervals when the X_l's are distributed equidistantly. Substituting $p = 1$ in the expression for the variance, gives the variance obtained in Expression (2). The limiting distribution of Y when $L \to \infty$, $n \to \infty$ and $n/L = \lambda$ remains constant, is the gamma distribution with shape parameter p and scale parameter $1/p\lambda$. A proof is given in Appendix B.2. The limiting distribution corresponds to the Z's being a stationary Poisson process on the real line, implying that the distance between every pth point follows a gamma distribution.

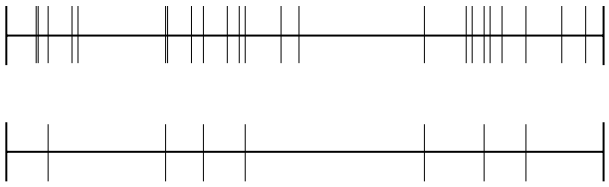


Figure 3: On the top line the $3n-1$ points are iid random variables with a uniform distribution $U[0, L]$. On the bottom line, only each 3rd ordered point is kept, giving $n-1$ points with repulsion between them.

6 Successive Partitioning of All Existing Intervals

The partitioning described in this section consists of successive partitioning of all existing intervals. The partitioning of the line is done in a stepwise procedure, where in each step all existing intervals are partitioned once. Thus, after k steps the number of intervals is $n = 2^k$. The position of the partitioning point on an existing interval is drawn uniformly on the interval. A partitioning of the interval $[0, L]$ is illustrated in Figure 4. The stepwise procedure is described in the following way:

Step 1 Let $X_1 \sim U[0, L]$, giving the interval lengths $Y_1^{(1)} = X_1$, $Y_2^{(1)} = L - X_1$, both having the distribution $Y_i^{(1)} \sim U[0, L]$.

Step 2 The two intervals with lengths $Y_1^{(1)}$ and $Y_2^{(1)}$ are partitioned by placing a point Z_i in each of them, where the points Z_i are uniformly distributed on the existing intervals:

- Let $Z_1|y_1^{(1)} \sim U[0, y_1^{(1)}]$, giving the interval lengths $Y_1^{(2)} = Z_1$, $Y_2^{(2)} = Y_1^{(1)} - Z_1$, both having the distribution $Y_i^{(2)}|y_1^{(1)} \sim U[0, y_1^{(1)}]$, $i = 1, 2$.
- Let $Z_2|y_2^{(1)} \sim U[0, y_2^{(1)}]$, giving the interval lengths $Y_3^{(2)} = Z_2$, $Y_4^{(2)} = Y_2^{(1)} - Z_2$, both having the distribution $Y_i^{(2)}|y_2^{(1)} \sim U[0, y_2^{(1)}]$, $i = 3, 4$.
- The points X_2 and X_3 are given by $X_2 = Z_1$ and $X_3 = X_1 + Z_2$.

Step 3 In the same way as in step 2, partition the existing intervals by drawing a point Z_i uniformly in each of the intervals. This yields $Y_i^{(3)}|y_j^{(2)} \sim U[0, y_j^{(2)}]$, where $i = 2j - 1, 2j$ and $j = 1, 2, 3, 4$. The points X_l, $l = 4, \ldots, 7$ can be found based on X_l, $l = 1, 2, 3$ and $Y_i^{(3)}$.

Step k After $k - 1$ steps there are 2^{k-1} intervals. Partition each of the intervals to get new intervals with length distributions $Y_i^{(k)}|y_j^{(k-1)} \sim U[0, y_j^{(k-1)}]$, where $i = 2j - 1, 2j$ and $j = 1, 2, \ldots, 2^{k-1}$. As above, the points X_l can be found.

Suppose the interval $[0, L]$ is partitioned into $n = 2^k$ intervals by the above procedure. The density function of an interval length $Y^{(k)}$ after k steps is then given by

$$f_{Y^{(k)}}(y) = \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L}\right)^{k-1}, \quad 0 \le y \le L. \tag{4}$$

The proof is given in Appendix A.3. The density function corresponds to the variable $W^{(k)} = -\ln(Y^{(k)}/L)$ following a gamma distribution, $W^{(k)} \sim \text{Gamma}(k, 1)$. The mean and variance of an interval length $Y^{(k)}$ are

$$\mathrm{E}\left[Y^{(k)}\right] = \frac{L}{2^k} = \frac{L}{n}, \qquad \mathrm{Var}\left[Y^{(k)}\right] = L^2\left(\frac{1}{3^k} - \frac{1}{2^{2k}}\right) = L^2\left(\frac{1}{n^{\ln 3/\ln 2}} - \frac{1}{n^2}\right).$$

Once again the mean equals the lengths of the intervals when the points X_l are equidistantly distributed.

7 Successive Partitioning of One Existing Interval

Siegel and Sugihara (1983) describe a sequential breakage procedure where at each stage in the procedure one of the existing intervals is further partitioned into two intervals. This

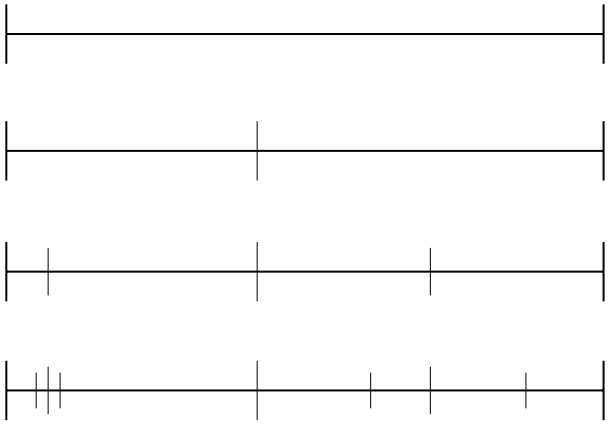


Figure 4: The 2^{k-1} existing intervals are partitioned into 2^k intervals. The partition points are drawn uniformly on each of the 2^{k-1} existing intervals.

partitioning is called a Kolmogorov breakage process. Suppose there are $n-1$ intervals at one stage. A move to the next stage is done by choosing one of the n intervals at random, with equal probability of each interval, and partitioning the interval according to a breakage distribution F. The resulting number of intervals is then n. Figure 5 shows a line partitioned in this way. For a partitioning following this procedure, the distribution of the resulting interval lengths will differ from the distribution described in Section 6, where all intervals are partitioned simultaneously.

Let $U^{(n)} = \log Y^{(n)}$ where $Y^{(n)}$ is the length of an interval when the total number of intervals is n. Siegel and Sugihara (1983) study the first two moments of the logarithm $U^{(n)}$. Using a uniform breakage distribution F, their results can be compared to the density function discussed in Section 6, where all intervals were partitioned in each stage. When F is the uniform distribution, the mean and variance of $U^{(n)}$ are

$$\mathrm{E}\left[U^{(n)}\right] = \ln L - 2\sum_{j=2}^{n}\frac{1}{j} \quad \text{and} \quad \mathrm{Var}\left[U^{(n)}\right] = 4\sum_{j=2}^{n}\frac{j-1}{j^2}, \tag{5}$$

where the variance of $U^{(n)}$ is independent of the length L. The density function of $Y^{(k)}$ in Expression (4) corresponds to the variable $W^{(k)} = \ln L - \ln Y^{(k)}$ following a gamma

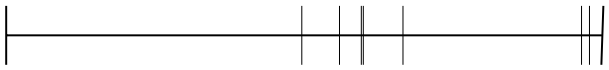


Figure 5: At each stage one interval is chosen at random and partitioned, with the partition point drawn uniformly on the chosen interval. All existing intervals have equal probability of being chosen.

distribution, $W^{(k)} \sim \text{Gamma}(k,1)$. Let $U^{(k)} = \ln Y^{(k)} = \ln L - W^{(k)}$, such that $(\ln L - U^{(k)}) \sim \text{Gamma}(k,1)$. The mean and variance of $U^{(k)}$ in this case are

$$\text{E}\left[U^{(k)}\right] = \ln L - k \quad \text{and} \quad \text{Var}\left[U^{(k)}\right] = k, \tag{6}$$

which can be compared to the mean and variance given in (5), where $n = 2^k$. Figure 6 shows plots of the different means and variances for $n \in [1, 100]$ and $L = 1$. The mean and variance given in (6) are both smaller in absolute value than the mean and variance given in (5).

Siegel and Sugihara (1983) study the two first moments of the logarithm $U = \ln Y$, but not the density functions of U or Y. Consider a uniform breakage distribution F. Theoretically, the density function of the interval lengths Y can be found, based on the results in Section 6. When there are n intervals, each interval length Y will have a length distribution with the density function given in (4), but the number k will differ for each interval, taking values in the set $\{1, \ldots, n-1\}$. Thus, conditioned on k the distribution of Y, $f_{Y|K}(y|k)$, is known. Let k_i be the number of k corresponding to the interval Y_i. All possible combinations of $(k_1, \ldots, k_n)$ can be found, and from the combinations the density function of K, $f_K(k)$, can be found. The conditional distribution of Y is then given by $f_Y(y) = \sum_k f_{Y|K}(y|k) f_K(k)$, and thus the density function of an interval length Y can be found. This is a simple task when n is small, but can be tedious work for large n, since there are $(n-1)!$ possible combinations of $(k_1, \ldots, k_n)$. Some of these combinations may be equal, but they result from different partitioning sequences. For $n = 4$ there are $3! = 6$ possible combinations of $(k_1, \ldots, k_4)$, and the density function of Y, expressed by the density functions $f_{Y^{(k)}}(y)$ in Expression (4), is

$$f_Y(y) = \frac{1}{6}\left(f_{Y^{(1)}}(y) + 3f_{Y^{(2)}}(y) + 2f_{Y^{(3)}}(y)\right) = \frac{1}{6}\left(1 - 3\ln y + (\ln y)^2\right), \quad 0 \le y \le 1.$$

For $n = 8$ there are $7! = 5040$ possible combinations of $(k_1, \ldots, k_8)$, and the density function of Y is a linear combination of $f_{Y^{(k)}}(y)$, $k = 1, \ldots, 7$.

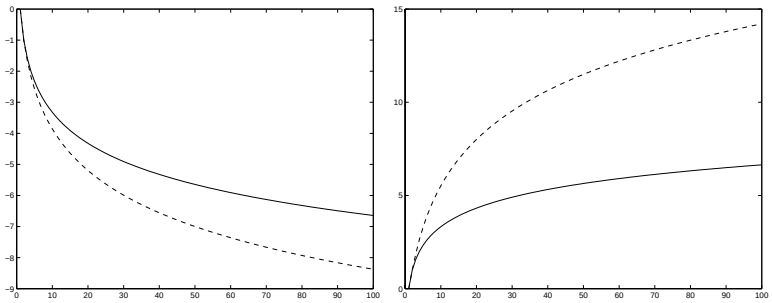


Figure 6: To the left is the mean of U plotted as a function of n, and to the right the variance of U. The dotted lines are the mean and variance given in Expression (5) and the solid lines are the mean and variance given in (6), where $k = \ln n / \ln 2$.

8 Comparisons of Length Distributions

Figures 7, 8 and 9 show the density functions for the interval lengths Y with different ways of partitioning. The density functions are shown for partitionings into 2, 4, 8, 16, 32 and 64 intervals, using original intervals of length $L = 1$. In Figure 7 the points X_l are independent uniformly distributed on $[0, L]$, in Figure 8 the intervals are created with repulsion between the interval endpoints, and in Figure 9 the intervals are created by successive partitioning of all existing intervals. In the case of independent uniformly distributed points the density function is monotonely decreasing, with $f_Y(0) = (n-1)/L$. The density function of the interval lengths produced by the successive partitioning of all existing intervals is also monotonely decreasing, but with $f_Y(y) \to \infty$ as $y \to 0$. When there is repulsion between the points, the density function is a unimodal function starting in $f_Y(0) = 0$. The mode of the function is at the point $y_{top} = L(p-1)/(pn-2)$, which approaches zero as $n \to \infty$.

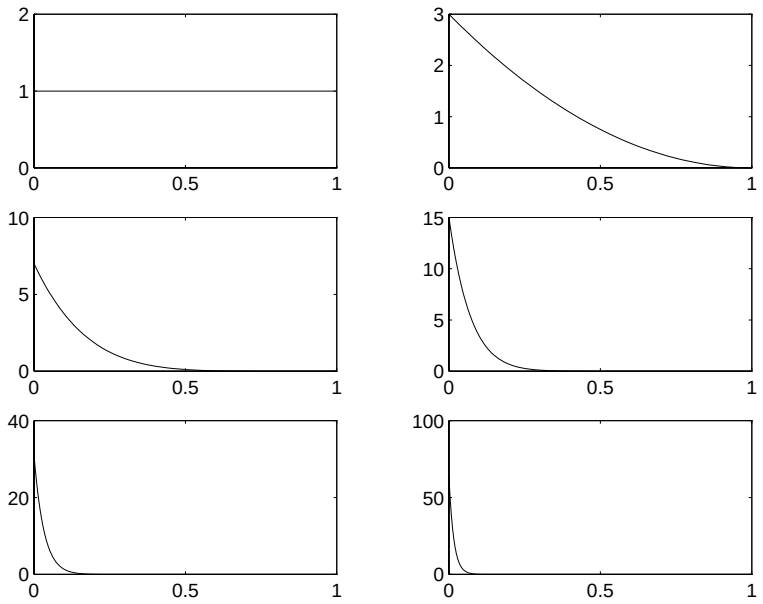


Figure 7: Density function of length distributions when the points X_l are independent uniformly distributed. The number of intervals is $2, 4$ (top), $8, 16$ (middle) and 32 and 64 (bottom).

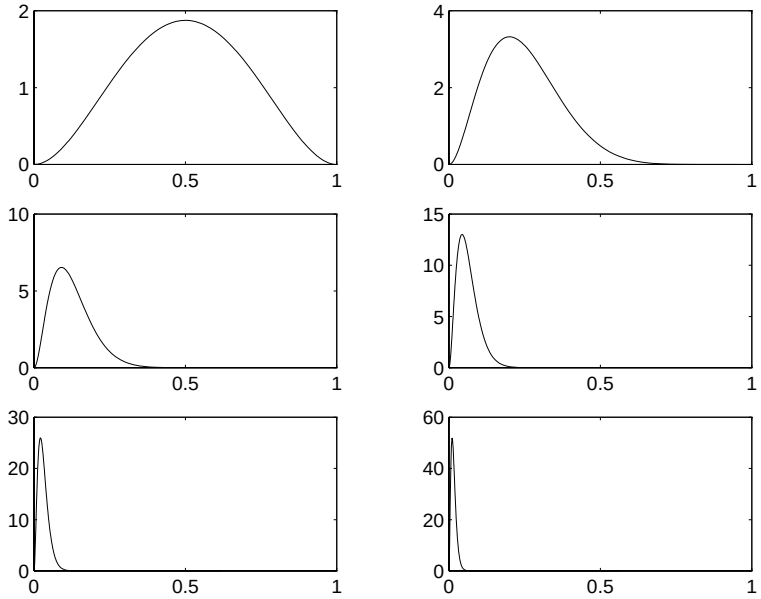


Figure 8: Density function of length distributions when the points X_l are given by $X_l = Z_{(3l)}$, where the points Z_j are independent uniformly distributed. The number of intervals is $2, 4$ (top), $8, 16$ (middle) and 32 and 64 (bottom).

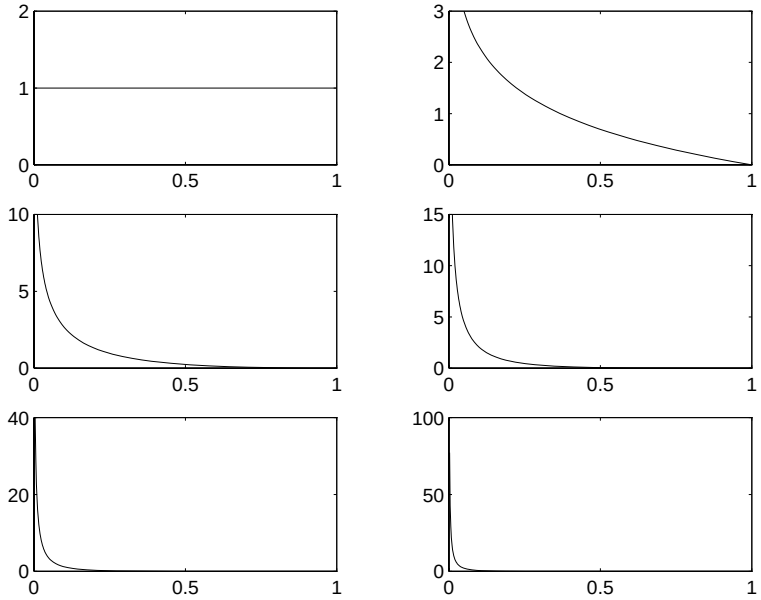


Figure 9: Density function of length distributions when the intervals are created by successive partitioning of all existing intervals. The number of intervals is $2, 4$ (top), $8, 16$ (middle) and 32 and 64 (bottom), which corresponds to $k = 1, \ldots, 6$.

9 Conclusions

Different models for partitioning of line segments have been discussed. In Section 3 the partitioning points are regularly spaced, resulting in equal lengths of all smaller intervals.

In Section 4 the partitioning points are independent identically uniformly distributed on the line. This is a special case of the model presented in Section 5, where points are independent identically uniformly distributed, but only every pth point is kept as a partitioning point, where p is a positive integer. The model in Section 5 enforce repulsion between the partitioning points when $p > 1$. The partitioning schemes in Section 4 and 5 results in beta distributions of the shorter line segments, when scaled by the length of the initial segment. The limiting distributions are studied, as the length of the initial interval and the number of partitionings approach infinity, while the relation between the length and number remains constant. The limiting distributions correspond to the initial points being a stationary Poisson process on the real line (Stoyan et al., 1995).

In Sections 6 and 7 sequential breakage processes are discussed. In Section 6 all existing intervals are partitioned simultaneously, while in Section 7 only one existing interval is partitioned in each step. The two sequential procedures lead to different distributions of the resulting intervals. The density function of the partitioning procedure described in Section 7 can, under certain conditions, be found based on the density functions found in Section 6.

References

Mouland, L. (1997). Gjennkjenning av multiple hendelser i tidsserier, Report in subject 75800 Prosjekt 2, Norwegian University of Science and Technology.

Siegel, A. F. and Sugihara, G. (1983). Moments of particle size distributions under sequential breakage with applications to species abundance, *Journal of Applied Probability* **20**: 158–164.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, New York.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2 edn, Wiley, New York.

A Computation of Length Distributions

A.1 Independent uniformly distributed partitioning points

This proof is a special case of the proof given in Appendix A.2. Let $X_{(0)} = 0$ and $X_{(n)} = L$. The interval lengths are then $Y_i = X_{(i)} - X_{(i-1)}$, $i = 1, 2, \ldots, n$. When the points X_l are independent uniformly distributed on the interval $[0, L]$, the density function of the lth ordered point $X_{(l)}$ is

$$f_{X_{(l)}}(x_{(l)}) = \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(l-1)!(n-l-1)!} \cdot x_{(l)}^{l-1}(L - x_{(l)})^{n-l-1}, \tag{7}$$

and the joint density of the neighbor points $X_{(l-1)}$ and $X_{(l)}$ is

$$f_{X_{(l-1)},X_{(l)}}(x_{(l-1)}, x_{(l)}) = \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(l-2)!(n-1-l)!} \cdot x_{(l-1)}^{l-2}(L - x_{(l)})^{n-1-l},$$

for $l = 2, \ldots, n-1$. The distribution of Y_i can be found by a bivariate transformation. Let

$$\begin{array}{ll} Y_i = X_{(i)} - X_{(i-1)}, & 0 \le Y_i \le L, \\ W = X_{(i)}, & Y_i \le W \le L, \end{array}$$

implying

$$\begin{array}{ll} X_{(i-1)} = W - Y_i, & 0 \le X_{(i-1)} \le L, \\ X_{(i)} = W, & X_{(i-1)} \le X_{(i)} \le L, \end{array}$$

with the Jacobian

$$J = \begin{vmatrix} \partial x_{(i-1)}/\partial y_i & \partial x_{(i-1)}/\partial w \\ \partial x_{(i)}/\partial y_i & \partial x_{(i)}/\partial w \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1.$$

The joint density of Y_i and W is then

$$\begin{aligned} f_{Y_i,W}(y, w) &= f_{X_{(i-1)},X_{(i)}}(w - y, w)|J| \\ &= \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(i-2)!(n-1-i)!} \cdot (w-y)^{i-2}(L-w)^{n-1-i}, \quad i = 2, 3, \ldots, n-1, \end{aligned}$$

and the marginal density of Y_i, $i = 2, 3, \ldots, n-1$, is found by

$$\begin{aligned} f_{Y_i}(y) &= \int_y^L f_{Y_i,W}(y, w)dw \\ &= \int_y^L \frac{1}{L^{n-1}} \cdot \frac{(n-1)!}{(i-2)!(n-1-i)!} \cdot (w-y)^{i-2}(L-w)^{n-1-i}dw \\ &= \frac{n-1}{L} \cdot \left(1 - \frac{y}{L}\right)^{n-2}. \end{aligned} \tag{8}$$

Since $Y_1 = X_{(1)}$ and $Y_n = L - X_{(n-1)}$, the density functions of Y_1 and Y_n are found from Expression (7), and are easily shown to equal the density function (8).

A.2 Repulsion between the partitioning points

This proof is similar to the proof in Appendix A.1, where the special case $p = 1$ is considered. We now find the distribution for general p. Let $Z_{(0)} = 0$ and $Z_{(pn)} = L$. The interval lengths are then $Y_i = X_{(i)} - X_{(i-1)} = Z_{(pi)} - Z_{(p(i-1))}$, $i = 1, 2, \ldots, n$. When the points Z_l are independent uniformly distributed on the interval $[0, L]$, the density function of the plth ordered point $Z_{(pl)}$ is

$$f_{Z_{(pl)}}(z_{(pl)}) = \frac{1}{L^{pn-1}} \cdot \frac{(pn-1)!}{(pl-1)!(p(n-l)-1)!} \cdot z_{(pl)}^{pl-1}(L - z_{(pl)})^{p(n-l)-1}, \tag{9}$$

and the joint density of the neighbor points $Z_{(p(l-1))}$ and $Z_{(pl)}$ is

$$\begin{aligned} & f_{Z_{(p(l-1))},Z_{(pl)}}(z_{(p(l-1))}, z_{(pl)}) \\ &= C(l) \cdot z_{(p(l-1))}^{p(l-1)-1}(z_{(pl)} - z_{(p(l-1))})^{p-1}(L - z_{(pl)})^{p(n-l)-1}, \quad l = 2, 3, \ldots, n-1. \end{aligned}$$

where

$$C(l) = \frac{1}{L^{pn-1}} \cdot \frac{(pn-1)!}{(p(l-1)-1)!(p-1)!(p(n-l)-1)!}.$$

The distribution of Y_i can be found by a bivariate transformation. Let

$$\begin{array}{ll} Y_i = Z_{(pi)} - Z_{(p(i-1))}, & 0 \le Y_i \le L, \\ W = Z_{(pi)}, & Y_i \le W \le L, \end{array}$$

implying

$$\begin{array}{ll} Z_{(p(i-1))} = W - Y_i, & 0 \le Z_{(i-1)} \le L, \\ Z_{(pi)} = W, & Z_{(i-1)} \le Z_{(i)} \le L, \end{array}$$

with the Jacobian

$$J = -1.$$

The joint density of Y_i and W is then

$$\begin{aligned} f_{Y_i,W}(y, w) &= f_{Z_{(p(i-1))},Z_{(pi)}}(w - y, w)|J| \\ &= C(i) \cdot (w - y)^{p(i-1)-1} y^{p-1}(L - w)^{p(n-i)-1}, \quad i = 2, 3, \ldots, n-1, \end{aligned}$$

and the marginal density of Y_i, $i = 2, 3, \ldots, n-1$, is found by

$$\begin{aligned} f_{Y_i}(y) &= \int_y^L f_{Y_i,W}(y, w)dw \\ &= \int_y^L C(i) \cdot (w - y)^{p(i-1)-1} y^{p-1}(L - w)^{p(n-i)-1} dw \\ &= \frac{p}{L} \cdot \binom{pn-1}{p} \left(\frac{y}{L}\right)^{p-1} \left(1 - \frac{y}{L}\right)^{p(n-1)-1}. \end{aligned} \tag{10}$$

Since $Y_1 = Z_{(p)}$ and $Y_n = L - Z_{(p(n-1))}$, the density functions of Y_1 and Y_n are found from Expression (9), and are easily shown to equal the density function (10).

A.3 Successive partitioning of all existing intervals

A partitioning of the line into $n = 2^k$ intervals is done by successive partitioning of all existing intervals. After the first partitioning the number of intervals is 2, after the second the number is $2^2 = 4$, and after $k-1$ steps the number of intervals is 2^{k-1}. The length distributions of the intervals created in step k, conditioned on the intervals after $k-1$ steps, are

$$Y_i^{(k)} | y_j^{(k-1)} \sim U[0, y_j^{(k-1)}], \quad i = 2j-1, 2j, \ j = 1, 2, \ldots, 2^{k-1}.$$

The density function of the length of the interval $Y_{2j}^{(k)}$ in the final partitioning is then given by the integral

$$f_{Y_{2j}^{(k)}}(y) = \int_y^L f_{Y_{2j}^{(k)}|Y_j^{(k-1)}}(y|y') f_{Y_j^{(k-1)}}(y') dy' = \int_y^L \frac{f_{Y_j^{(k-1)}}(y')}{y'} dy'. \tag{11}$$

The density function of $Y_{2j-1}^{(k)}$ equals that of $Y_{2j}^{(k)}$. Since $Y_j^{(1)} \sim U[0, L]$, $j = 1, 2$, $f_{Y_j^{(1)}}(y)$ is known, and $f_{Y_j^{(k)}}(y)$, $k = 2, 3, \ldots$, $j = 1, 2, \ldots, 2^k$ can be found. The density function is identical for all $j = 1, 2, \ldots, 2^k$, and is given by

$$f_{Y^{(k)}}(y) = \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L}\right)^{k-1}, \quad 0 \le y \le L. \tag{12}$$

This is shown by induction on k. If $k = 1$, the interval length $Y^{(1)}$ is uniformly distributed on the interval $[0, L]$, giving

$$f_{Y^{(1)}}(y) = \frac{1}{L},$$

which equals Expression (12) with $k = 1$. If $k = 2$ the integral in (11) can be solved to give

$$f_{Y^{(2)}}(y) = \int_y^L \frac{f_{Y^{(1)}}(y')}{y'} dy' = \int_y^L \frac{1}{y'L} dy' = \frac{1}{L} \cdot \left(-\ln \frac{y}{L}\right),$$

which equals Expression (12) with $k = 2$. Thus Expression (12) is valid for $k = 1, 2$. Suppose (12) holds for $k-1$. Using the integral in (11) the density function $f_{Y^{(k)}}(y)$ is found to be

$$\begin{aligned} f_{Y^{(k)}}(y) &= \int_y^L \frac{f_{Y^{(k-1)}}(y')}{y'} dy' = \int_y^L \frac{1}{y'} \cdot \frac{1}{L(k-2)!} \cdot \left(-\ln \frac{y'}{L}\right)^{k-2} dy' \\ &= \frac{1}{L(k-1)!} \cdot \left(-\ln \frac{y}{L}\right)^{k-1}. \end{aligned}$$

B Limiting Distributions

B.1 Independent uniformly distributed partitioning points

The density function of Y, given in Expression (1), corresponds to $W = Y/L$ following a beta distribution, $W \sim \text{Beta}(1, n-1)$. The moment generating function of Y can be found through the moment generating function of W:

$$M_Y(t) = \text{E}\left[\exp(tY)\right] = \text{E}\left[\exp(tLW)\right] = M_W(tL).$$

Using the moment generating function of the beta distribution, we get

$$\begin{aligned} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \left\{ \prod_{r=0}^{k-1} \frac{1+r}{n+r} \right\} \cdot \frac{(tL)^k}{k!} = 1 + \sum_{k=1}^{\infty} \frac{k!}{n \cdot (n+1) \cdots (n+k-1)} \cdot \frac{(tL)^k}{k!} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(tL)^k}{n \cdot (n+1) \cdots (n+k-1)} = 1 + \sum_{k=1}^{\infty} \frac{t^k}{\frac{n}{L} \cdot \left(\frac{n}{L} + \frac{1}{L}\right) \cdots \left(\frac{n}{L} + \frac{k-1}{L}\right)} \\ &= 1 + \sum_{k=1}^{\infty} \frac{t^k}{\lambda \cdot \left(\lambda + \frac{1}{L}\right) \cdots \left(\lambda + \frac{k-1}{L}\right)}. \end{aligned} \tag{13}$$

In the last step we let $n/L = \lambda$. For $t < \lambda$, taking limits of the sum in (13) gives

$$\lim_{L \to \infty} M_Y(t) = 1 + \sum_{k=1}^{\infty} \left(\frac{t}{\lambda}\right)^k = \sum_{k=0}^{\infty} \left(\frac{t}{\lambda}\right)^k = \frac{1}{1 - \frac{t}{\lambda}} = \frac{\lambda}{\lambda - t},$$

which is the moment generating function of the exponential distribution with parameter λ.

B.2 Repulsion between the partitioning points

The density function of Y, given in Expression (3), corresponds to $W = Y/L$ following a beta distribution, $W \sim \text{Beta}(p, p(n-1))$. As in Appendix B.1, the moment generating function of the beta distribution can be used to find the limiting distribution of Y as $L \to \infty$, $n \to \infty$ and $n/L = \lambda$. The proof is similar to the proof in Appendix B.1, where the special case $p = 1$ is considered. For general p we find

$$\begin{aligned} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \left\{ \prod_{r=0}^{k-1} \frac{p+r}{pn+r} \right\} \cdot \frac{(tL)^k}{k!} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(p+k-1)!}{(p-1)!k!} \cdot \frac{t^k}{\frac{pn}{L} \cdot \left(\frac{pn}{L} + \frac{1}{L}\right) \cdots \left(\frac{pn}{L} + \frac{k-1}{L}\right)} \\ &= 1 + \sum_{k=1}^{\infty} \binom{p+k-1}{k} \cdot \frac{t^k}{p\lambda \cdot \left(p\lambda + \frac{1}{L}\right) \cdots \left(p\lambda + \frac{k-1}{L}\right)}. \end{aligned} \tag{14}$$

For $t < \lambda$, taking limits of the sum in (14) gives

$$\begin{aligned}\lim_{L\to\infty} M_Y(t) &= 1 + \sum_{k=1}^{\infty} \binom{p+k-1}{k} \left(\frac{t}{p\lambda}\right)^k = \sum_{k=0}^{\infty} \binom{p+k-1}{k} \left(\frac{t}{p\lambda}\right)^k \\ &= \left(\frac{1}{1-t/p\lambda}\right)^p = \left(\frac{p\lambda}{p\lambda - t}\right)^p,\end{aligned}$$

which is the moment generating function of the gamma distribution with shape parameter p and scale parameter $1/p\lambda$.

Appendix A

Model choice for fault distribution

Model choice for fault distribution

Hilde G. Borgos and Henning Omre,
Department of mathematical sciences, Norwegian University of Science and Technology (NTNU),
N-7034 Trondheim, Norway. Email: hgb@stat.ntnu.no, omre@stat.ntnu.no.

Key words

Fault distribution, Power law, Bayes factor, Markov chain Monte Carlo sampling (MCMC).

Abstract

The objective of the presentation is to check the validity of the power law model for fault size. The approach is to establish a plausible competing model, and to compare the two. The comparison is done according to statistical tradition by using Bayes factors.

Geologists often use a power law to express the relationship between fault size and cumulative number:

$$n = bx^{-a}$$

where x is either the maximum displacement or the length of the fault, and n is the cumulative number of faults larger than x. To define a fault density function for x corresponding to this power law, a lower limit of the fault size must be chosen, in order for the density function to exist. If the fault sizes follow the power law, a plot of cumulative number versus fault size on log-log scale should produce a straight line of negative slope a.

However, real data usually give a log-log plot of cumulative number versus fault size with a concave shape. We will suggest another density function for the fault distribution. This distribution does not follow the power law, but will give a relation between size and cumulative number producing a concave curve plotted on log-log scale. Furthermore, the density function will be defined for any x larger than zero, hence there is no need for a lower limit on the fault size.

We will compare the fit of the two models on a given data set. To do so, we estimate the parameters in each model, and the Bayes factor. The Bayes factor gives an indication of which model fits the data best.

The approach can be done as: Let θ_i denote the parameters under model i, $i = 1, 2$, and let x be the observed fault sizes. The likelihood of x under model i, given the parameters θ_i is expressed through the likelihood function $f_i(x|\theta_i)$. If we specify prior distributions $f_i(\theta_i)$ for the model parameters, the posterior distributions for the parameters are

$$f_i(\theta_i|x) \propto f_i(\theta_i) f_i(x|\theta_i).$$

This can be used for estimating the model parameters by use of MCMC sampling. The Bayes factor is the ratio between the marginal densities of x under model 1 and model 2, defined as:

$$B(x) = \frac{f_1(x)}{f_2(x)},$$

where

$$f_i(x) = \int f_i(\theta_i) f_i(x|\theta_i) d\theta_i.$$

The Bayes factor can also be estimated in the MCMC sampling procedure.

Hilde G. Borgos will present the paper.

Appendix B

Stochastic simulation of fault patterns conditioned on seismic data and well data

Stochastic Simulation of Fault Patterns Conditioned on Seismic Data and Well Data

Hilde Borgos
Norwegian University of Science and Technology, Department of Mathematical Sciences, NTNU, Trondheim, N-7034, Norway

Sedimentary rocks tend to be rather homogeneous laterally, and have a non-homogeneous vertically layered structure. This heterogeneity structure is reflected in the seismic observations, where lateral discontinuities visible offsetting the vertical layers are interpreted as geological faults. However, the interpretation of faults from seismic maps contain uncertainties, both due to the quality of the seismic data and to the judgment of the interpreter.

In this work the problem of detecting faults from 3D seismic data is considered from a statistical point of view. Stochastic modeling is used in an attempt to detect faults which displace sedimentary layering from 3D seismic observations and well observations. The lateral coherence observed in seismic data is incorporated in the model, and discontinuities in this coherency are used to locate faults. Observations from wells are also integrated in the model, and constitute additional information about the reservoir.

A Bayesian framework is used to present a stochastic model for fault patterns, and to condition on available information like seismic data and well observations. Let Δ be a stochastic variable representing the fault pattern. General geological knowledge about faults and fault patterns is described by a prior distribution $f(\delta)$ of the variable Δ. Let R denote the available observations of the reservoir. The observations are integrated in the model through the likelihood function $f(r|\delta)$, giving the likelihood of observing R if the true fault pattern is Δ. Combining the prior distribution and the likelihood function, the posterior distribution is obtained according to Bayes rule:

$$f(\delta|r) = \text{const} \times f(\delta) f(r|\delta)$$

The posterior distribution of the fault pattern Δ, conditioned on the observations R, contain both general geological knowledge and reservoir specific observations. By sampling from the posterior distribution $f(\delta|r)$, realizations of the fault pattern Δ are obtained. These realizations should reflect the discontinuities present in the observations. A number of realizations of fault patterns are generated to study the variability of the posterior distribution. Markov chain Monte Carlo techniques are used to sample from the posterior distribution.

The fault pattern Δ is represented by a hexagonal grid image. The pixel values P give the topography of a horizon resulting from the faulting of originally horizontal layers. The fault planes are assumed to be vertical, and all faults are assumed to penetrate completely through the vertical extent of the region under study. Thus the vertical displacements of the faults are large compared to the vertical range of the observations. The edges E are used to draw fault traces on the horizon, and thus define the fault planes. Fault traces are composed by connecting adjacent grid points. The vertices V (the grid points) represent the vertical displacement along the fault traces, and vertical displacement values are assigned to all vertices along a trace. The sets E, V and P are treated as stochastic variables, and the fault pattern Δ is given as a function of these variables; $\Delta = \Delta(E, V, P)$. The prior distribution of Δ is factorized as $f(\delta) = f(e, v, p) = f(e) f(v|e) f(p|e, v)$. Most focus will be put on the prior distribution $f(e)$ of the edges E, representing the pattern of fault traces. The prior distribution is based on the work by Tjelmeland and Besag (1998), and is represented by a Markov Random Field (MRF):

$$f(e) = \text{const} \times \exp\{-U(e)\}$$

where $U(e)$ is an *energy function*. A low energy $U(e)$ corresponds to a high probability of the pattern of fault traces produced by e. The energy function is defined through a sum of *potentials*, where different configurations of edge values are assigned different potentials. Thus the appearance of the fault traces can be controlled through the definitions of the different potentials, making some characteristics of the traces more likely than others. For example, certain fault directions or intersection angles can be favored. Two examples of realizations of E, using different potentials, are shown in Figure 1. The vertical displacement values along a fault trace are modeled as a 1D Gaussian Random Field, and a MRF conditioned on edge and vertex values is used to model the pixel values and thus the topography of the faulted horizon.

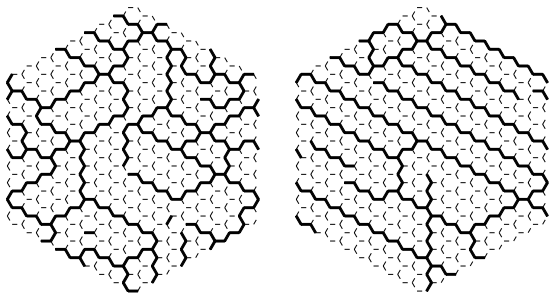


Figure 1: *Realizations from the prior distribution of fault traces.*

The likelihood function $f(r|\delta)$ is based on the work by Eide *et al.* (1997). Available observations R consist of seismic data and well observations. Based on well logs, the reflection coefficients in the wells are obtained. In some situations the reflection coefficients of a sedimentary rock, before faulting, can be modeled as a 3D Gaussian Random Field (GRF). Furthermore, the relationship between reflection coefficients and seismic amplitudes can be modeled as a linear convolution, thus the joint distribution of seismic amplitudes and reflection coefficients are represented also by a GRF. After faulting the original structure of the rock is altered. However, if all displacements due to faulting are known, the observations of the reservoir can be transformed back to obtain the original structure. After this transformation is done, the transformed observations can be modeled as a GRF. The likelihood function $f(r|\delta)$ is given as the probability density function of a multivariate Gaussian distribution, where the observations R are transformed according to the fault pattern Δ. The Gaussian likelihood function is analytically tractable and is easily calculated even for relatively large datasets.

In the presentation further details of the model will be described. Examples of realizations sampled from the posterior distribution will be shown, to illustrate how fault patterns can be generated via conditioning on seismic data and well data.

Acknowledgments

I would like to thank Håkon Tjelmeland for his assistance on this work.
The research is funded by VISTA.

References

Eide, A. L., Ursin, B. and Omre, H. (1997). Stochastic simulation of porosity and acoustic impedance conditioned to seismic data and well data, *67th Annual Meeting of the Society of Exploration Geophysicists, Expanded Abstracts*, SEG, Dallas, Texas.

Tjelmeland, H. and Besag, J. (1998). Markov Random Fields with Higher-Order Interactions, *Scandinavian Journal of Statistics* **25**(3): 415–433.

Appendix C

Uncertainty in fault geometries

Uncertainty in Fault Geometries

Hilde G. Borgos and Henning Omre
Department of Mathematical Sciences, Norwegian University of Science and Technology,
Trondheim, Norway. Email: hgb@stat.ntnu.no, omre@stat.ntnu.no
Ref.: http://www.math.ntnu.no

Abstract

A stochastic model of fault geometries is constructed, enabling fault patterns to be conditioned on available observations from a petroleum reservoir, consisting of three dimensional stacked seismic observations and reflection coefficients obtained from well logs. A Gaussian model is used for the observations, where fault geometries are visible as irregularities in the Gaussian field. The model for fault geometries is defined in a Bayesian framework, where the likelihood model of observations is incorporated together with prior general geological knowledge about fault patterns. Repeated sampling from the model is used to study the uncertainty in fault geometries.

INTRODUCTION

Stochastic modeling of petroleum reservoirs can be used to forecast production, and to assess the associated uncertainty. A model for reservoir characteristics is first defined, based on available observations. A forecast of production is then obtained through a numerical simulation of fluid flow, using the specified recovery procedure. The aim of this work is to define a stochastic model for the fault pattern in a reservoir. The number of faults, their sizes and locations have influence on fluid flow and on the connectivity of the reservoir, and thus on the production forecast.

Two types of information about the fault pattern in a reservoir are available. General geological information about fault characteristics is included in the model. Reservoir specific information provide data on the presence and locations of faults. For example, from seismic observations large faults are detected as discontinuities in the lateral continuity in seismic amplitudes, where the discontinuities are visible throughout a number of subsequent horizontal layers. The stochastic model of fault patterns is constructed in a Bayesian framework, enabling fault patterns to be conditioned on available observations like seismic observations and observations from wells. The uncertainty in fault geometries can be studied through repeated sampling from the posterior stochastic model.

Earlier work on stochastic modeling of faults or fractures include object based techniques, see for example Chilès (1988), and pixel or grid based approaches, e.g., Gringarten (1996). The cited papers present models where observations are used to estimate parameters related to characteristics of fault geometries, while no likelihood function is defined to describe the relationship between fault realizations and available observations. Construction of non-stochastic algorithms for tracking faults from seismic observations was presented by Bahorich and Farmer (1995, 1996), introducing the coherence cube. The coherence cube is a moving window algorithm calculating the correlation between neighbor traces of seismic signals, using local information to reveal discontinuities. The current approach is significantly different from the mentioned publications. A formal stochastic model conditioning on a larger set of seismic traces and well observations is used to assess the uncertainty in the fault pattern.

BAYESIAN FRAMEWORK AND BASIC ASSUMPTIONS

A stochastic model is defined for a faulted horizon H, where the fault pattern is visible as discontinuities in the surface. The aim of this work is to generate geological fault patterns by sampling of the faulted horizon H, conditioned on the available observations, $\boldsymbol{O}$. A Bayesian framework is used to define the relationship between the faulted horizon and the observations. A prior model $f(h)$ of the horizon includes general geological knowledge about faults and fault patterns. The relationship between the faulted horizon and the observations are defined through the likelihood function $f(\boldsymbol{o}|h)$. Combining the prior distribution and the likelihood function, the posterior distribution is obtained,

$$f(h|\boldsymbol{o}) = \text{const} \times f(\boldsymbol{o}|h)\, f(h). \tag{1}$$

The posterior distribution combines general geological knowledge and reservoir specific observations. Assessment of uncertainty in fault geometries is obtained by repeated sampling from the posterior distribution.

In the construction of the stochastic model, some basic assumptions are made. The vertical extents of the faults are assumed to be large compared to the region under study, thus all faults penetrate completely through the vertical extent of the region. Furthermore a crude approximation is made, assuming vertical faults with constant offset within this region.

PRIOR MODEL OF GEOLOGICAL FAULT PATTERNS

In this work post sedimentary faults are considered, and the sedimentary layers of rock are assumed to be originally horizontal. Due to faulting rock is displaced, with a spatial variation in the magnitude of the displacement. As a result, an originally horizontal surface is deformed into a topographic surface in three dimensions, referred to as a horizon. Faults are identified as discontinuities in the horizon. Other natural variations like ductile deformation, folding and tilting are assumed to produce small displacements of the rock compared to the faulting, and result in continuous variations in the altitude of the horizon. Under the assumption of faults penetrating completely through the region under study, the topography of subsequent horizons are comparable. The fault pattern can then be represented by a single horizon H which is given as a topographic map of altitudes $H = H(x, y)$, where the faults, their locations, orientations and offsets, are recognized from the deformations on the surface caused by faulting.

The prior model $f(h)$ of the faulted horizon H includes information of typical characteristics observed in geological fault patterns. For example, information of intersection angles, which are typically close to 60° or 30°, anisotropies in fault orientations and displacement profiles along fault traces are incorporated in the prior model. The faulted horizon is represented by a stochastic gray-scale image based on a hexagonal grid, where both edges, vertices and pixels contain information about the surface, see Figure 1. The hexagonal grid is preferred since it simplifies modeling of intersection angles of 60°. The quantities included in the prior model of a faulted horizon are fault traces on the horizon, offset profiles along the fault traces and the overall topography of the surface.

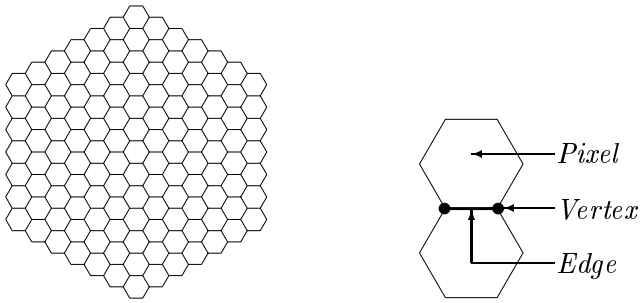


Figure 1: (a) Hexagonal grid used in the image representation of a faulted horizon. (b) Illustration of an edge (E), vertex (V) and pixel (P) in the hexagonal grid.

The edges $\boldsymbol{E} = (E_1, E_2, \ldots, E_{n_e})$ are binary variables with values $E_i \in \{0, 1\}$, and by connecting edges of value $E_i = 1$ fault traces on the horizon are obtained. Vertices $\boldsymbol{V} = (V_1, V_2, \ldots, V_{n_v})$ are used to model the offsets along the fault traces, their magnitudes and directions, while

vertices off the fault traces are assigned zero value. The pixels $\boldsymbol{P} = (P_1, P_2, \ldots, P_{n_p})$ give the altitude of the horizon H at the center point of each pixel. Altitudes are given in relative coordinates, and does not give an absolute depth. Based on the discrete image representation, the continuous surface $H = H(\boldsymbol{E}, \boldsymbol{V}, \boldsymbol{P})$ is found deterministically by triangulation, accounting for the discontinuities along fault traces. The prior model of the faulted horizon H can be factorized as

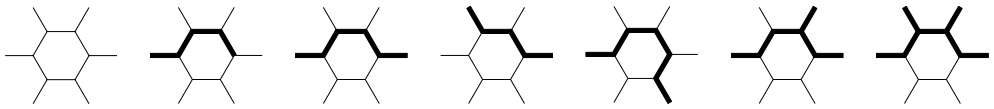
$$f(h) = f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v})\, f(\boldsymbol{v}|\boldsymbol{e})\, f(\boldsymbol{e}) \tag{2}$$

where $f(\boldsymbol{e})$, $f(\boldsymbol{v}|\boldsymbol{e})$ and $f(\boldsymbol{p}|\boldsymbol{e}, \boldsymbol{v})$ are prior distributions of edges, vertices and pixels.

A Markov random field is used as a prior distribution of edges, $f(\boldsymbol{e})$. According to the Hammersley-Clifford theorem, see for example Winkler (1995), the distribution of any Markov random field $\boldsymbol{E}$ can be expressed as a Gibbs distribution (Geman and Geman, 1984) with a density function of the form

$$f(\boldsymbol{e}) = \text{const} \times \exp\{-\sum_{c \in \mathcal{C}} \omega_c(\boldsymbol{e}_c)\}. \tag{3}$$

The set $\mathcal{C}$ is the set of all cliques, where a clique is defined as a set of neighboring edges, and $\omega_c(\boldsymbol{e}_c)$ is the potential function of the edges $\boldsymbol{e}_c$ contained in the clique c. The choice of potential functions are inspired by the work of Tjelmeland and Besag (1998). All possible configurations of edges within cliques of size 12 are sorted into specified classes. Figure 2 shows some examples: background, end points, lines, angles and intersections. By assigning different potentials to each class, specific configurations of fault traces are favored. The configurations can also be sorted by direction, giving higher probability to some orientations of fault traces.



(a) (b) (c) (d) (e) (f) (g)

Figure 2: The figure shows some of the classes of clique configurations defined in the prior distribution of edges. (a) A clique where all edges have value $E_i = 0$ is classified as background. Other classes are (b) end point, (c)–(d) lines, (e) angle and (f)–(g) single and double intersection.

The prior model of vertices given edges, $f(\boldsymbol{v}|\boldsymbol{e})$, is a simple model where deterministic offset profiles are assigned to each fault trace. Symmetric offset profiles with a maximum offset near the center of the trace and zero offset at the fault tips are used for isolated fault traces, see for example Dawers, Anders and Scholz (1993). For non-isolated fault traces with fault tips terminating in other fault traces, modifications of the offset profile are made. The number of fault traces and their lengths are determined by the edge values $\boldsymbol{E}$. In our model, only the

direction of the offset is modeled as a stochastic variable, with two possible directions for each fault trace.

A multivariate Gaussian distribution is used as a prior model for the pixels given edges and vertices, $f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v})$. The mean vector μ accounts for discontinuities along fault traces. The positions of the fault traces and the magnitude of the offsets depend on the values of $\boldsymbol{E}$ and $\boldsymbol{V}$, thus the mean value μ is given as a function of the edge and vertex values. The prior distribution of pixels is

$$f(\boldsymbol{p}|\boldsymbol{e},\boldsymbol{v}) = \phi(\boldsymbol{p};\mu(\boldsymbol{e},\boldsymbol{v}),\Sigma_p) \tag{4}$$

where $\phi(\boldsymbol{p};\mu,\Sigma)$ is the pdf of the multivariate Gaussian distribution with mean vector μ and covariance matrix Σ. Note that the final offset values along fault traces may differ slightly from the vertex values $\boldsymbol{V}$, and are determined by the difference in pixel values at opposite sides of the fault traces.

LIKELIHOOD MODEL OF RESERVOIR SPECIFIC OBSERVATIONS

Available observations $\boldsymbol{O}$ are included in the model through the likelihood function $f(\boldsymbol{o}|h)$, giving the likelihood of the observations conditioned on the faulted horizon H. Available observations from the reservoir consist of three dimensional seismic data and well logs. An underlying model is defined for reflection coefficients $\boldsymbol{C}$ in three dimensions, and a convolution model is used to model the relationship between reflection coefficients and seismic amplitudes. The observed seismic signals $\boldsymbol{S}^o$ can be expressed as a linear combination of reflection coefficients $\boldsymbol{C}$ and independent Gaussian noise $\boldsymbol{U} \sim N(0,\sigma_s^2 I)$,

$$\boldsymbol{S}^o = A(\psi_s)\boldsymbol{C} + \boldsymbol{U} \tag{5}$$

where $A(\psi_s)$ is a convolution matrix defined by the seismic wavelet with wavelet parameter ψ_s. Based on the well logs, exact observations of reflection coefficients $\boldsymbol{C}^o$ can be obtained at a limited number of locations.

Consider first a sedimentary reservoir prior to any faulting. Reflection coefficients of the reservoir can then be modeled as a Gaussian random field (Eide, Ursin and Omre, 1997; Eide, 1999). From Expression (5) it follows that the seismic signals will then be a Gaussian random field, and the joint distribution of seismic signals and reflection coefficients obtained from well logs can be expressed as a multivariate Gaussian distribution

$$\boldsymbol{O}' \sim N(\mu_o',\Sigma_o') \tag{6}$$

where $\boldsymbol{O}'$ are the observations that would be obtained if gathered prior to faulting. After faulting the original structure of the rock is altered. Recall the model assumptions discussed above, of vertical faults with constant offset and penetrating completely through the vertical extent of the region under study. The fault pattern is represented through the faulted horizon $H(x, y)$. At a point (x, y) the reflection coefficients originally located in (x, y, t) is then displaced to the point $(x, y, t + H(x, y))$, see Figures 3a and 3b. If the faulted horizon $H(x, y)$ is known, observations obtained from the reservoir can be transformed back to their original location, reproducing the initial structure. This is illustrated in Figure 3c.

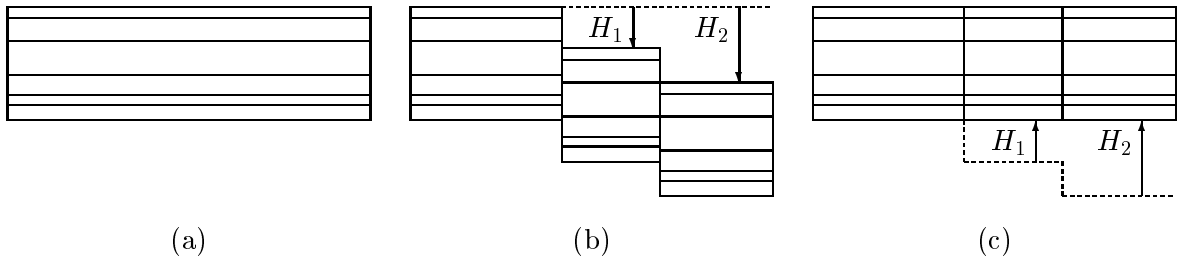


Figure 3: Figure (a) illustrates the structure of the rock prior to faulting. After faulting the structure is altered, as in figure (b). If the faulted horizon H is known, a transformation of observations back to the original structure can be performed, as shown in figure (c).

The position of an observation C^o or S^o gathered from a faulted reservoir does not coincide with the original position prior to faulting. The marginal distributions of the observations are unchanged, but due to the spatial rearrangement the mean vector μ'_o and covariance matrix Σ'_o are no longer valid. The transformation of locations can be expressed as a permutation of the elements of the vector $\boldsymbol{O}'$, giving the observations $\boldsymbol{O}$ of the faulted reservoir, and a corresponding rearrangement of the order of the elements in the mean vector and covariance matrix. The overall distribution of the observations $\boldsymbol{O} = (\boldsymbol{S}^o, \boldsymbol{C}^o)$ is then a Gaussian distribution,

$$[\boldsymbol{O}|H = h] \sim N(\mu_o(h), \Sigma_o(h)) \tag{7}$$

where $\mu_o(h)$ and $\Sigma_o(h)$ are obtained from μ'_o and Σ'_o by rearranging the positions of the elements according to the transformation dictated by the faulted horizon $H = h$. The likelihood function of the observations $\boldsymbol{O}$ is thus the pdf of a multivariate Gaussian distribution:

$$f(\boldsymbol{o}|h) = \phi(\boldsymbol{o}; \mu_o(h), \Sigma_o(h)). \tag{8}$$

SAMPLING FROM THE POSTERIOR MODEL

A posterior distribution $f(h|\boldsymbol{o})$ of the faulted horizon conditioned on available observations is obtained by combining the prior model and the likelihood function. Markov chain Monte Carlo (McMC) techniques are used to generate realizations from the posterior model, see Besag et al.

(1995) and references therein for an introduction to McMC theory. The realizations from the posterior distribution should reflect the discontinuities present in the observations. Not only the position of the discontinuities are located, but also the offsets of the faults. By generating numerous realizations from the posterior model, the uncertainty of the fault pattern in the posterior model can be studied.

The McMC algorithm used to sample from the posterior distribution involves evaluation of the Gaussian likelihood function in Expression (8). The theoretical expression of the likelihood is well known, and can be calculated if the dimension of the observations is small. As the dimension increases however, the likelihood can in practice not be obtained analytically for general Gaussian fields. An approximation to the likelihood function must be found, which in this work consists of two approximations. First, the likelihood function of the observations $\boldsymbol{O}$ is approximated by a pseudo-likelihood function

$$f(\boldsymbol{o}|h) \approx \prod_i f(\boldsymbol{o}_i|\boldsymbol{o}_{-i}, h) \tag{9}$$

where $\boldsymbol{O}_i$ is chosen as a vertical column through the observations, and $\boldsymbol{O}_{-i}$ is the set of all other observations. Next, the conditional distribution is approximated as

$$f(\boldsymbol{o}_i|\boldsymbol{o}_{-i}, h) \approx f(\boldsymbol{o}_i|\boldsymbol{o}_{\partial i}, h) \tag{10}$$

where only observations $\boldsymbol{O}_{\partial i}$ in a neighborhood ∂i of observations $\boldsymbol{O}_i$ are included in the conditioning.

SIMULATION EXAMPLES

The performance of the model presented above is illustrated through simulation examples based on synthetic observations. A three dimensional Gaussian field on a rectangular grid of size $50 \times 50 \times 20$ is first generated, representing the reflection coefficients of the reservoir prior to faulting. Corresponding seismic observations are obtained using the convolution model in Expression (5). Next, a fault pattern and a corresponding faulted horizon are generated from the prior model in Expression (2), see Figure 4, and the resulting structure of the faulted reservoir is found by a rearrangement of the seismic observations, see Figure 5. Hence observations $\boldsymbol{O} = \boldsymbol{S}^o$ of a faulted reservoir are obtained, and samples of the faulted horizon can be generated conditioned on the observations. For these synthetic observations, the true fault pattern and faulted horizon are known, see Figure 4, thus a comparison can be made with the conditional samples.

The hexagonal image representation of the fault pattern covers the lateral extent of the observations, and only the square part of the image where observations are available is illustrated in Figure 4. The image of edges in Figure 4a represents the pattern of fault traces on the surface. The horizon in Figure 4b is discontinuous along the fault traces, with a smooth variation in altitude between faults. The main feature of the fault pattern is two parallel north-west striking faults of opposite dip directions. Figure 4c shows a horizontal cross section through the seismic observations at vertical position $t = 13$, where the faults can be detected as discontinuities in the seismic signals. Figure 5 shows a north-south oriented vertical cross section through the seismic observations at position $x = 25$. Thus the cross section is obtained through the center part of the observations, and crosses the two major faults.

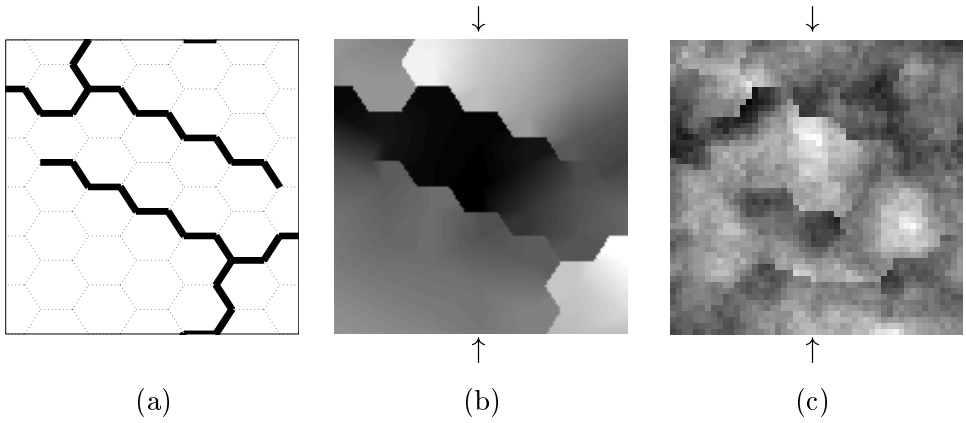


Figure 4: (a) Pattern of fault traces $\boldsymbol{E}$ and (b) faulted horizon H in the square area of observations. The two north-west striking faults are regarded as straight lines, the jagged appearance is an artifact of the hexagonal grid. Figure (c) shows a horizontal cross section through the seismic observations $\boldsymbol{S}^o$, at vertical position $t = 13$, see Figure 5. The arrows mark the position $x = 25$ of the vertical cross section illustrated in Figure 5.

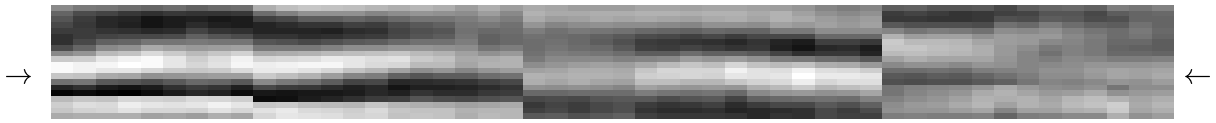


Figure 5: North-south oriented vertical cross section through the seismic observations $\boldsymbol{S}^o$ at position $x = 25$, see Figure 4. The arrows mark the position $t = 13$ of the horizontal cross section in figure 4c.

Figure 6 shows three realizations of the faulted horizon H from the posterior distribution $f(h|\boldsymbol{o})$, based on the synthetic observations described above. The realizations are obtained by performing 50,000 iterations of the McMC sampling algorithm. The algorithm is restarted each time, thus there is no dependency between the realizations. All realizations reproduce the main features of the two dominating faults, while the region near the fault tips differ between the realizations.

Figure 7 shows cross sections of observations corresponding to the cross section in Figure 5, but with observations transformed back to original positions according to the generated faulted horizons in Figure 6. The concept of this transformation is illustrated in Figure 3. After the transformation, the discontinuities produced by the faults should be removed and the original strong horizontal coherence should be retrieved. The fault with the largest offset, to the right in Figure 5, is no longer detectable in the transformed observations based on the two first realizations. In the third realization the offset of this fault is slightly under-represented.

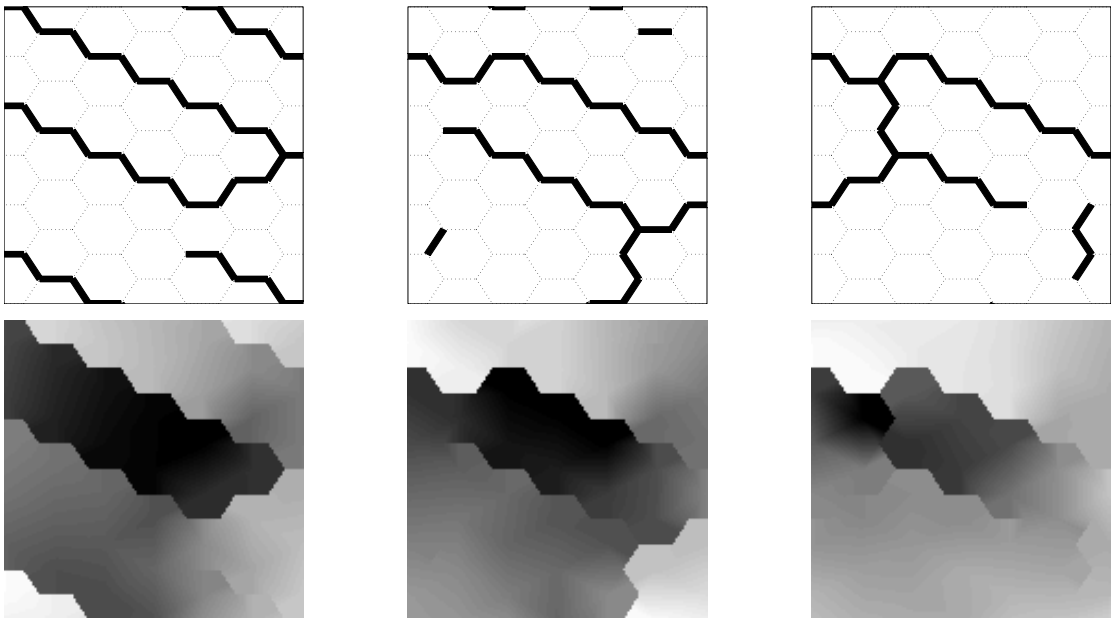


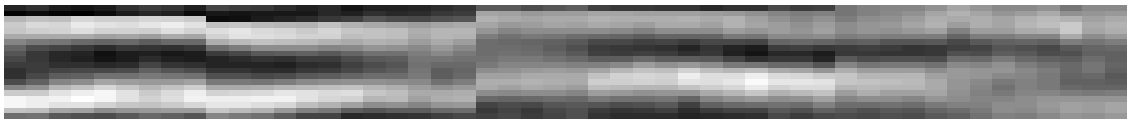
Figure 6: Realizations of pattern of fault traces $\boldsymbol{E}$ (top) and faulted horizon H (bottom) generated from the posterior distribution $f(h|\boldsymbol{o})$ of fault patterns.

CLOSING REMARKS

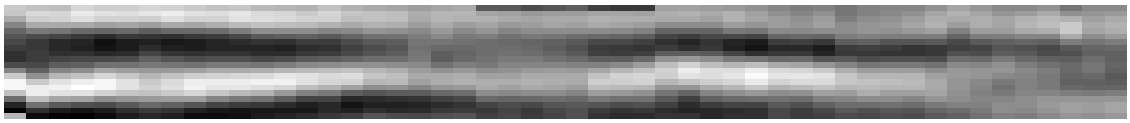
The paper presents a consistent stochastic model for fault geometries, conditioned to seismic data and well observations, and defined in a Bayesian framework. Through a synthetic data example, it is demonstrated how the stochastic model can be used to assess uncertainties in fault geometry at seismic scales.

Acknowledgments

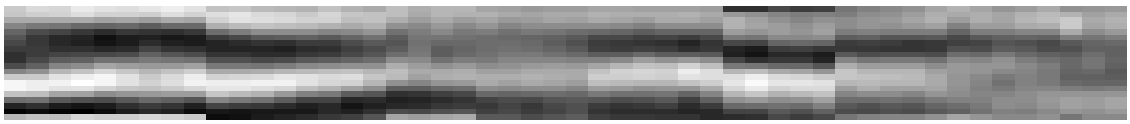
The authors are grateful to Chris Townsend for giving a basic understanding of the geological aspects related to this problem. We also thank Håkon Tjelmeland for his assistance on this work. H. G. Borgos is funded by a PhD grant from VISTA (*Det Norske Videnskaps-Akademi* and *Statoil*).



Realization 1



Realization 2



Realization 3

Figure 7: Cross sections corresponding to the section in Figure 5, but with observations transformed according to the generated faulted horizons in Figure 6.

References

Bahorich, M. S. and Farmer, S. L. (1995). The coherence cube, *The Leading Edge* **14**(10): 1053–1058.

Bahorich, M. S. and Farmer, S. L. (1996). Method of seismic signal processing and exploration. U. S. Patent Number 5,563,949.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion), *Statistical Science* **10**(1): 3–66.

Chilès, J. P. (1988). Fractal and Geostatistical Methods for Modeling of a Fracture Network, *Mathematical Geology* **20**(6): 631–654.

Dawers, N. H., Anders, M. H. and Scholz, C. H. (1993). Growth of normal faults: Displacement-length scaling, *Geology* **21**: 1107–1110.

Eide, A. L. (1999). *Stochastic Reservoir Characterization Constrained by Seismic Data*, PhD thesis, Norwegian University of Science and Technology. Doktoringeniøravhandling 1999:5.

Eide, A. L., Ursin, B. and Omre, H. (1997). Stochastic simulation of porosity and acoustic impedance conditioned to seismic data and well data, *67th Annual Meeting of the Society of Exploration Geophysicists, Expanded Abstracts*, SEG, Dallas, Texas.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6): 721–741.

Gringarten, E. (1996). 3-D Geometric Description of Fractured Reservoirs, *Mathematical Geology* **28**(7): 881–893.

Tjelmeland, H. and Besag, J. (1998). Markov Random Fields with Higher-Order Interactions, *Scandinavian Journal of Statistics* **25**(3): 415–433.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer.