Quantification and assessment of fault uncertainty and risk using stochastic conditional simulations

Li Shuxing, Roussos Dimitrakopoulos

(W H Bryan Mining Geology Research Centre, the University of Queensland, Brisbane Qld 4072, Australia)

Abstract  The effect of geological uncertainty on the development and mining of underground coal deposits is a key issue for longwall mining, as the presence of faults generates substantial monetary losses. This paper develops a method for the conditional simulation of fault systems and uses the method to quantify and assess fault uncertainty. The method is based on the statistical modelling of fault attributes and the simulation of the locations of the centres of the fault traces. Fault locations are generated from the thinning of a Poisson process using a spatially correlated probability field. The proposed algorithm for simulating fault traces takes into account soft data such as geological interpretations and geomechanical data. The simulations generate realisations of fault populations that reproduce observed faults, honour the statistics of the fault attributes, and respect the constraints of soft data, providing the means to thereby model and assess the related fault uncertainty.

Keywords  fault systems, longwall mining, quantification of fault uncertainty and risk, simulation, fractal model

Introduction

Longwall mining may suffer heavy financial losses through the presence of faults that generate delays in production schedules, changes in mine plans, and loss of coal reserves. Mining operations will be improved if methods are developed to quantify the uncertainty of faults, assess geological interpretations and integrate such quantitative information in longwall design and planning and production management. A key issue is the ability to quantitatively model the uncertainties associated with faults and quantify the associated risk.

Research to date has been inadequate in the areas of understanding fault uncertainty and simulating fault systems in longwall coal mining. Hatherly et al. (1993) present a study on fault prediction based on structural and geophysical methods. Fault simulation in longwall mining is shown in Li et al. (1999) and Dimitrakopoulos and Li (2000), and is founded on the inhomogeneous Poisson process, an approach limited in taking spatial correlations in fault locations and soft information into account. Theoretical developments and applications of stochastic simulation of faults are available in the petroleum industry, where interest is focused on the simulation of sub-seismic

* Supported by the Australian Coal Research Association for selecting and funding this research project (ACARP C7025); Anglo Coal Australia, Newlands Coal Mine for additional funding and collaboration; and BHP Coal, CSIRO, North Goonyella Mine and MIM Holdings for collaboration.

E-mail: shuxing.li@uq.edu.au
faults, joints and small fractures. Recent developments include marked point processes and factorial kriging (Wen and Sinding-Larsen, 1997) and Gibbs fields (Omre et al., 1994; Munthe et al., 1993).

In this paper, fault systems are simulated with a new approach which enables uncertainty in longwall mining to be quantified. The approach is based on the use of correlated probability values for the thinning of a Poisson process which is used to generate the locations of the centres of the fault traces. Probability fields are simulated in an indicator framework and accommodate soft data coded as prior distributions. The paper describes the algorithm for stochastic conditional simulation of fault systems. The application of the algorithm in a coalfield follows. Subsequently, the quantification of fault uncertainty is presented and ramifications for longwall mining are discussed.

1 Fractal models for fault characterisation

The application of fractal theory to faults can be traced back to the 1980s and includes the work of King (1983), Turcotte (1986) and Childs et al. (1990). In general, the theory suggests that various fault parameters are invariant with respect to scale or are 'self similar', providing a model that can be used for predictive purposes. Gauthier and Lake (1993) provide a succinct outline of the principles and characteristics of fractal fault models.

The use of fractal or self-similar models in studying and characterising fault systems is widespread in the technical literature. During the 1990s a consensus developed that power-law scaling models describe size-frequency relationships and the relationship between the length and maximum displacement (throw) of fault populations, despite the discussion and different opinions on how these properties are inferred and interpreted (e.g. Walsh et al., 1994; Nicol et al., 1996; Needham and Yielding, 1996; Marret et al., 1999; Bour and Davy, 1999). Although limited, Doutsos and Koukouvelas (1998) and Watterson et al. (1996) present data from studies in coal seams that also support the same fractal relations.

1.1 Power-law models of fault size distributions

In fractal theory, fault size (throw or length) distributions may be described by the following power-law (fractal) model over a wide range of fault size such that

$$\log(N_s) = \alpha - \beta \log(S),$$

where, $N_s$ is the cumulative number of faults with either length or throw greater than size $S$; $S$ is the fault size (length $L$ or throw $T$); $\alpha$ is a scaling factor that is a function of the fault density in a given area (when $\alpha$ is high, the fault density is high); and $\beta$ is the fractal dimension of the fault population that defines the relative number of large and small faults (when $\beta$ is high, the number of small faults is high relative to the number of large faults). For a given $\beta$, the power-law model of the fault size distribution indicates how the larger faults and smaller faults are proportionally distributed in a fault population. This property enhances understanding of a fault population and may predict missing faults from a set of known ones.

Fig. 1 shows a typical fault size (throw or length) distribution (cumulative number of faults versus maximum fault throw) from a seismic data set. The plot shows: (1) The data forming a typical "central segment" used to model the fractal throw relation; (2) A flattened upper segment re-
fleeing truncated or under-sampled faults with small fault displacements (sub-seismic faults); (3) A steep lower segment due to the incomplete representation or censoring of the larger faults of the population which are expected to exist outside the area limits of the seismic survey.

In addition, the figure shows the position on the graph of data from the core scale for the same population which are expected to fit the same power-law model.

1.2 Power-law models for length-throw relationships

The inter-relationship between fault length and maximum throw is widely accepted as following a power-law written as

$$T_{\text{max}} = cL^n,$$

where, $T_{\text{max}}$ is the maximum fault throw; $L$ is the fault length; $c$ is a constant reflecting rock properties; and $n$ is a constant. Both $c$ and $n$ are obtained from data and are subject to ambiguities in fault interpretations. For practical purposes in previous investigations, and again in this study, the centre of the fault is considered to be the location of maximum displacement.

Fig. 2 shows the plot of a typical maximum throw—versus—length relationship. The relationship between fault throw and length is very often subject to substantial scatter for a number of reasons, particularly sampling effects. More specifically, the scatter is often attributed to difficulties in discerning the true end of a fault, to the extension of faults beyond the boundaries of the survey and to part of the fault length being below the limit of resolution of the survey technology. The underestimation of small fault lengths is typical when faults are sampled from maps or along traverses or in cross-sections (Marrett and Allmendinger, 1991). The throw component also contributes to the observed scatter; when seismic surveying is used, the resolution limit results in a relatively greater underestimation of the throws of smaller throw faults than of larger throws. Maximum displacement is not always recorded. The throw—length diagrams often span only one order of magnitude and together these factors may generate biases. The scatter of the data and the resulting diversity of $n$ values in Equation (2) may also arise as a consequence of geological events. Reactivation events can lead to more diverse $n$ values by complicating the identification of fractal populations, and fault linkage leads to a highly subjective sampling of branched fault lengths.

2 Fault simulation based on probability fields

The algorithm proposed herein is based on the statistical description of a fault population as described in the previous section and the thinning of a Poisson process (Stoyan and Stoyan, 1994) using a correlated probability field similar to the probability fields proposed by Srivastava (1992). The proposed conditional simulation of faults proceeds as follows (Fig. 3):

(1) Define a random path to be followed in visiting locations $x$, which are to be considered as centres of fault traces. There are $N$ locations or grid nodes $\{x_i, i = 1, \ldots, N\}$ to be potentially visited. The $N$ locations exclude the known fault centres.

(2) Generate a realisation of an auto-correlated probability field $\{p(x_i), i = 1, \ldots, N\}$ reproducing the uniform marginal cumulative distribution function (cdf) and the covariance corresponding to the covariance of the uniform transform of the fault densities in the study area.

(3) Estimate at the first location $x_i$, the intensity function of an inhomogeneous Poisson process $\lambda(x_i)$ using a planar Epanechnikov kernel estimator.
Use the probability value \( p(x_i) \) at location \( x_i \) to thin a Poisson process from
\[
1 - p(x_i) < \lambda(x_i) / \lambda^*,
\]
where, \( \lambda^* \) is the intensity of a corresponding homogeneous Poisson process and \( \lambda(x) \leq \lambda^* \). If the above constraint is met, a fault centre exists at \( x_i \), If not, the next node on the random path is visited until the constraint is met.

Select a maximum fault throw from the power-law model of the fault size distribution in Equation (1).

Grow the fault from the centre of a fault trace, in opposite directions, using the following steps: ① Define the fault strike by randomly drawing from the fault strike distribution. ② Define the fault length from the power-law model of the relationship between fault length and maximum throw in Equation (2). ③ Use a distance step and a directional tolerance at each step to grow the fault. ④ Stop the growth when the fault length has been reached.

Repeat steps (3) to (6) until the number of total faults satisfies the fault size distribution in Equation (1).

Repeat the process to generate additional realisations.

The fault simulations described above reproduce the faults identified within a study area, and existing and simulated faults reproduce the desired statistical fault characteristics.

The simulation of the probability field in Step (2) of the above algorithm requires further consideration. The probability field \( p(x_i) \) can be generated by either a Gaussian or indicator non-conditional simulation algorithm. An indicator sequential simulation algorithm (Alabert, 1987) facilitates the use of multiple secondary information and it is implemented for this study.

The integration of soft information into fault simulation by means of the simulation of the probability field has a major advantage of incorporating all types of information relevant to fault generation and spatial distribution. The soft information is coded as prior probability, which includes the rock type of the roof and floor, the thickness of the competent roof and floor, the area extent of the competent roof and floor, the major folding structure, the information from remote-sensing and the observations of expert geologists. This will enable a robust model for fault simulation to be built and the fault uncertainty to be quantified.

3 Simulating fault systems in a coalfield

3.1 Data description and statistics

Fig. 4 (a) shows the study area where the development of potential underground coal mines is being investigated. Faults occur at a variety of scales within the study area and are generally either north–north–westerly trending thrust faults or north–easterly oriented normal faults. Thrust faults have throws ranging from \( 6 \) m to \( 25 \) m. Normal faults display highly variable displacements and dominate the mine area. There are a few folds, including gentle synclines, and sharper anticlines and monoclines. Folds trend north–north–west, fold axes plunge south and thrust blocks upthrow to the east.
Fault data used in this study are interpreted from drilling, surface mapping, 3D seismic surveys and studies of the regional structural geology. Faults are mapped out on a coal seam horizon and are spatially distributed as shown in Fig. 4 (a). The rose diagram of fault strikes is also shown in Fig. 4 (b).

The corresponding fault size (throw) power-law distribution is modelled using Equation (1) and is shown in Fig. 5 (a) (Population 1). The specific models inferred for Population 1 are \( \log (N_T) = 1.9927 - 1.103 \log (T) \) (pessimistic) and \( \log (N_T) = 1.8696 - 0.9576 \log (T) \) (optimistic). It is interesting to note that the data points at the top and bottom end of each curve in Fig. 5 (a) deviate from the model fitted. The top end deviation is due to small faults not identified by drilling, mapping, 3D seismic and geological interpretations. The bottom end deviation is due to the limited area covered during the geological study upon which the analysis is based.

The relationship between fault length and throw for Population 1 is shown in Fig. 5 (b). A power-law model is fitted according to Equation (2), giving \( T_{\text{max}} = 0.0245L^{0.8037} \) (pessimistic) and \( T_{\text{max}} = 0.08519L^{0.678} \) (optimistic).

The power-law model for Population 2 is shown in Fig. 6 (a) by \( \log (N_T) = 1.885 - 1.1019 \log (T) \) (pessimistic) and \( \log (N_T) = 1.8067 - 0.9877 \log (T) \) (optimistic). The relationship between fault length and throw for Population 2 is shown in Fig. 6 (b) by \( T_{\text{max}} = 0.02935L^{0.8378} \) (pessimistic) and \( T_{\text{max}} = 0.05648L^{0.7804} \) (optimistic).

The power-law model for Population 3 is shown in Fig. 7 (a) by \( \log (N_T) = 1.5682 - 0.9233 \log (T) \) (pessimistic) and \( \log (N_T) = 1.4858 - 0.7816 \log (T) \) (optimistic). The relationship between fault length and throw for Population 3 is shown in Fig. 7 (b) by \( T_{\text{max}} = 0.02766L^{0.8547} \) (pessimistic) and \( T_{\text{max}} = 0.05086L^{0.7422} \) (optimistic).

Because of the quality of the data used, a pessimistic model and an optimistic model are produced for the fault throw distribution and the relationship between fault length and throw for each population. In reality, it is difficult to obtain high quality fault data because the data are obtained either from a limited small area, or by insufficient
means. However, using a pessimistic model and an optimistic model to describe the fractal characteristics of a fault population can alleviate the data quality effect.

3.2 Estimation of fault density

The planar Epanecnikov kernel estimator for estimating point density in an inhomogeneous Poisson point process as described by Stoyan and Stoyan (1994) is

$$\lambda(x) = \frac{8}{3\pi h} \sum_{i=1}^{n} e_{k|x-x_i|},$$

$$e_{k|x-x_i|} = \begin{cases} 
\frac{3}{4h} \left(1 - \frac{|x-x_i|^2}{h^2}\right) & (|x-x_i| < h), \\
0 & \text{(otherwise)}
\end{cases}$$

where, $\lambda(x)$ is the density function; $x$ is the location where the density is estimated; $x_i$ is a location within distance $h$ from $x$; and $h$ is a distance termed ‘smoothing factor’.

A fault can be assumed to be a point in an inhomogeneous Poisson point process, that point being the central point of the fault trace. In practice the fault data obtained rarely cover the whole study area. Using the above planar Epanecnikov kernel estimator will present two problems; 1. If the smoothing factor $h$ is not large enough, some locations within the study area may become blank spots and have zero fault density, possibly due to no fault data being recorded in that part of the area. 2. If the smoothing factor $h$ is too large, it will smooth out the structural details of the fault distribution. It is hard to balance the degree of smoothness and the removal of the blank spots.
A varying neighbourhood point density estimation method is therefore proposed for the fault density estimation. As the name indicates, the method uses the same planar Epanechnikov kernel, but instead of a fixed neighbourhood (smoothing factor \( h \)) with a varying number of points within the neighbourhood, it uses a varying neighbourhood (smoothing factor \( h \)) with a fixed number of points within the neighbourhood. This effectively eliminates the blank spots in the study area. The smoothness of the density estimation can also be adjusted.

### 3.3 Dealing with soft information

Soft information integrates all geological information relevant to faulting other than known fault data. It reflects the geological environment of faulting. It is used in fault simulation as described in step (2) of the fault simulation algorithm. All types of soft information are coded into prior probability indicators, which are subsequently updated into posterior probabilities by indicator simulations.

In this study, the soft data used to divide the lease area into regions with a similar propensity for faulting are based on local indicators of faulting and some geotechnical information regarding the roof lithology. Monoclines and the regions of steep dips associated with them are good indicators of faulting within the area. Monocline zones have very high soft probabilities (Table 1) and are slightly better indicators of faulting than the steeply dipping parts of the area.

Massive channel sandstones overlying the coal seam also provide an indication of the locations of faults in the area, for these competent units tend to propagate faults at their boundaries and into weaker units. The channel sandstones trend north–east to south–west across the lease, and an area surrounding the sandstone was defined where the soft probability of faulting is high.

The spatial distribution of other lithological units in the roof strata provided further means of dividing the lease into areas with similar propensity for faulting. Roof strata were divided into three sections each with different geotechnical characteristics. Roof type A is characterised by interbedded sandstone and siltstone. Roof type B is similar to type A but is underlain by highly carbonaceous siltstone and mudstone. Roof type C is also similar to type A but is underlain by massive sandstone. Roof type C is the strongest and will therefore have a low soft probability, while Roof type B, the weakest, will have a high soft probability.

In combining all the soft information sources, it was necessary to decide which source is most likely to find previously unidentified faults. Local indicators of faulting and sandstone boundaries are the most predictive sources of soft information. In all other areas, the roof strata provide the best indication of possible faulting. Soft data sources are the same for all fault populations.

Average soft probabilities assigned to the sources discussed above are presented in Table 1. The probability assigned to each source is reported in the middle column. The probability of the intersection of local indicators with the sandstone boundary is given in the third column.

After the soft information is numerically coded (Li et al., 2001), it is used to generate soft probabilities of faulting for the simulation of the fault populations (Fig. 8). It is important to note that the area of the open pit was given a probability of ‘0’. It is assumed that all faults have been identified in this area, and in the area directly adjacent to the open pit. Note that the soft probability values shown in Table 1 reflect the relative variation of faulting probability between each zone and have been assigned to maintain the relative order of probability rather than to be an exact value.
Fig. 8 One realisation of probability field for three populations

The relative ranking of different parts of the study area provides prior probabilities that are not, technically, true probabilities (hence their description as ‘soft’). The prior soft values are updated using the method of sequential indicator simulation. Posterior probabilities are generated and these correspond to true probabilities. Underlying probability fields (P-fields), which are used to assess the selection of fault location, were generated for each fault population realisation based upon the geological information available and inputs to the method of sequential indicator simulation.

3.4 Conditional simulations

The fault simulation algorithm described in section 2 is used here to simulate the fault population in the study area. The simulation is based on the population statistics discussed in section 3.1 and the underlying spatial characteristics of the fault system. The underlying spatial characteristic of a fault system is accounted for by the modelling of continuity and anisotropy of the spatial fault distribution. The variogram of fault density reveals the direction of greater continuity and the anisotropy ratio. Fault density is estimated using the method described in section 3.2. Fig. 9 and Table 2 shows the variograms for Populations 1, 2 and 3 respectively.

Fig. 9 Variogram models using fault density data for Populations 1, 2 and 3

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The variograms for Populations 1, 2 and 3 respectively</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction</td>
<td>Type - P1</td>
</tr>
<tr>
<td></td>
<td>Sill</td>
</tr>
<tr>
<td>Nugget</td>
<td>0.06</td>
</tr>
<tr>
<td>Spherical</td>
<td>Southeast</td>
</tr>
<tr>
<td></td>
<td>Northeast</td>
</tr>
</tbody>
</table>
Based on the underlying characteristic of spatial fault distribution, and other geological information, the fault probability fields can be simulated. The geological information consists of lithological and geotechnical characteristics of seam roof strata, monoclinal zones and the distribution of massive channel sandstones in this study. Fig. 8 shows one probability field realisation for Populations 1, 2 and 3 from left to right. The simulated fault probability fields are used in deciding the locations of simulated faults along with the fault density information.

Distance step and direction tolerance used in step 3 of the simulation are 200 m and 5° respectively. Fig. 10 shows two realisations of fault simulation in the study area based on the three probability fields in Fig. 8. The realisations reproduce the original faults and their statistical characteristics including the power-law model of throw distribution, the model of the relationship between fault length and throw, and the distribution of fault strikes. At the same time the realisations also reproduce the underlying spatial characteristics of the fault population identified by the variogram modelling. Fig. 11 shows the reproduction of the strike for the fault populations in the area and the reproduction of the power-law models of the throw in the same realisation. The example suggests that the reproduction of data characteristics is excellent.

Fig. 10 Two realisations of all fault populations using the pessimistic models

Fig. 11 Pessimistic power-law models and strike distribution reproduced in simulations for Populations 1, 2 and 3

3.5 Fault uncertainty assessment and its application

A large number of fault realisations are combined to generate probability maps over the study area, showing the probability of having faults of given specifications of interest. Fig. 12 shows probability maps having faults with throws over 1 m in the case study. This probability mapping is a useful tool to measure the uncertainty of the fault system studied. In particular, the fault uncertainty information is spatially presented. This facilitates its application in longwall coal mining.

The fault probability maps quantify the uncertainty of having faults with throws larger than 1 m that can be used
to select low risk areas for mine development, such as the area in the northern part of the study area. Furthermore, the probabilities of faults may also be locally used to optimise the layout of longwall panels so as to minimise the fault occurrence within them. Fig. 13 shows the integration of the quantified fault uncertainty information into a longwall design and plan. Through the integration of quantified fault risk, a longwall panel layout may avoid the fault risk and be optimised.

Fig. 12 Fault probability map for all populations using the pessimistic models

Additional scenarios may be generated based on different geological interpretations, coding of prior information or the use of additional sources of soft data. Finally, the uncertainty of the fault statistics used may be assessed.

4 Conclusions

A method to simulate fault systems in two dimensions is founded upon the statistical characterisation of fault populations and the use of this characterisation to simulate equally probable, geologically meaningful fault populations. A new algorithm is based on the thinning of a Poisson process which is used to place the centres of fault traces and on generally accepted power-law models that define interrelationships of fault attributes. The thinning of a Poisson process is based on a simulated probability field reflecting the spatial continuity of fault densities and integrating soft data, including geological and geomechanical interpretations. The algorithm also honours the available faults at their locations and reproduces their statistics.

The combination of large numbers of realisations of fault systems in an area of interest for longwall coal mining provides probability maps that quantify the uncertainty of having a fault in a given location. Probability maps are tools that can enhance longwall design, as well as the planning and management of risk.

Quantification of geological risk provides practically useful results that can be incorporated into the decision-making process in a way that is not possible when using qualitative (i.e., subjective, interpretative or relative) risk assessments. Qualitative risk assessments are unable to be linked to mine design and planning, whereas quantitative risk assessments that incorporate local geological understanding offer accurate risk assessments and can be used directly in mine design and planning.

References


