Influence of Geology on Valley Upsidence and Closure

Mine Subsidence Engineering Consultants
Peter DeBono

David Shteinman MIEAust
Chief Operating Officer
ACCM

Australian Centre for Commercial Mathematics
University of New South Wales
Overview

• Our goal is to develop a probabilistic model for closure based on four explanatory variables:
  • valley depth
  • longitudinal distance from longwall
  • transverse distance from longwall
  • maximum subsidence

• We are particularly interested in predicting the likelihood and size of extreme closures – worst case scenarios (e.g. 1 in 1000 event – the closure size which has a 0.1% probability of being exceeded OR the 99.9% percentile).

• This poses a statistical challenge because we are interested in extreme deviations from the overwhelming majority of the available data.

• We have employed a branch of statistics known as Extreme Value Theory to overcome this challenge.
Extreme Value Theory tells us that the probabilities of extreme events are given approximately by a Generalized Pareto Distribution (GPD).

The GPD (red) may have a *heavy or fat tail* when compared to the Normal distribution (blue) – the tail is important when modelling rare events.
• We have built a model for the probability of extreme closures by fitting the Generalized Pareto distribution to our data.

• This involves fitting two parameters (or coefficients) – sigma (scale) and xi (shape) – to the data.

• We fit the parameters at many different values of the four explanatory variables (Valley Depth, Longitudinal Distance, Transverse Distance and Maximum Subsidence) – this gives us local models based on only a small range of the data.

• We then fit the two parameters as smooth functions of the four explanatory variables. This gives us an overall model over the full range of the four explanatory variables.
Example of parameter fitting:
Sigma v Longitudinal distance

The points on this graph are local fits for sigma (the scale parameter of the GPD) based on the data points close to a particular Longitudinal distance value. The Red line is the parametric model developed based on these values.
Model output gives a predicted 1 in 1000 event (red line)

The dots are MSEC data points for (longitudinal distance, closure). The red line is the 1 in 1000 return level for our parametric model.

Evaluate model by comparing predicted event occurrence to the data.
RESULTS

- Our parametric approach is data-driven and hence reproducible, given the dataset.

- The GPD probabilistic model is robust to the effects of outliers in the observed closure data.

- We account for the effects of all four explanatory variables simultaneously, which prevents possible prediction errors due to highly correlated explanatory variables.

- The model may be advanced by adding further explanatory variables and interactions between predictors.

- The probabilistic model can also provide confidence bands on our predicted 1 in 1000 returns, giving a measure of prediction accuracy.