Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Rock Mechanics & Mining Sciences 43 (2006) 454–472
International Journal of Rock Mechanics and Mining Sciences
www.elsevier.com/locate/ijrmms

Reliability of numerical modelling predictions

T.D. Wiles*

Mine Modelling (Pty) Limited, PO Box 637, Mt. Eliza, Vic., Australia 3930

Accepted 4 August 2005
Available online 28 September 2005

Abstract

The accuracy of all predictions made using numerical modelling is strictly limited by the natural variability of geologic materials. In this paper, an attempt is made to quantify this accuracy through the straightforward application of probability and statistics. It is shown how the contributions from variability of the input parameters and also errors introduced by the modelling procedure can be combined into a single representative coefficient of variation C_p. This parameter is a characteristic that quantifies how well the entire modelling procedure is performing. It includes contributions from the variability of the pre-mining stress and rock mass strength, material heterogeneity and also errors introduced by the modelling procedure (e.g. elastic versus inelastic), and represent the uncertainty one has in predictive capability.

In the paper, a lower limit for C_p of 30% is estimated for use with the conventional empirical approach (i.e. measurement of pre-mining in situ stress state, laboratory testing, and subsequent strength degradation to rock mass scale). Realistic values are most likely higher than this since some contributions have not been included and others are not known with any certainty. Various methods to reduce the magnitude of this parameter are then investigated. It is shown how this parameter can be evaluated by back-analysis of field observations. An example is detailed where a series of pillar failures are back analysed to calculate a site-specific value of 10%. This allows predictions to be made with greatly improved confidence and accuracy, and demonstrates why the back-analysis approach is so appealing. The paper presents a rational means for improving on existing empirical procedures for design of underground excavations.

Keywords: Numerical modelling; Prediction reliability; Mine design; Back-analysis; Empirical failure criterion

1. Introduction

It has long been recognized that geotechnical modelling problems are data limited. While it is not uncommon in the civil engineering environment to devote several percent of the project budget to rock mass characterization, in mining situations this figure is normally several orders of magnitude smaller. This necessitates a very different modelling approach from that developed in, e.g., civil, electrical or aerospace engineering [1]. The objective of this paper is to provide a methodology for determining quantitative accuracy limits and applying those in a meaningful, practical way to design problems. This paper is meant to take up the challenge posed by Hoek [2] to "find a better way" than conventional empirical procedures to design underground excavations.

Rock failure occurs when stresses exceed the rock mass strength. There is considerable natural variability in the in situ pre-mining stress and rock lithology, deformation properties and strength. This results in uncertainty in both the accuracy of stress predictions and the strength to which these are compared. In addition to this ambiguity, the numerical model used, whether elastic, inelastic or otherwise, always represents an approximation to the actual rock mass behaviour. Even the most complex material models still require simplifying assumptions to be made about one or more parameters. As a consequence the accuracy of all failure predictions using numerical modelling will be limited.

Geological materials are naturally very non-uniform. Practical considerations limit the amount of information that can be determined about the geology and behavioural

*Tel.: +61 3 9787 0870; fax: +61 3 9787 9008.
E-mail address: twiles@map3d.com.

1365-1609/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijrmms.2005.08.001

properties of the rock mass. Since real fracture system geometry and heterogeneity of the properties will never be known, numerical models can only represent a very small proportion of the system behaviour. For these reasons, numerical models can only simulate reality with limited accuracy. Owing to the irregularity of the rock mass, the real behaviour of the later cannot be known with absolute confidence. Practical considerations limit the amount of information that can be determined about the rock mass response. Hence calibration of numerical models can only be conducted with limited accuracy.

Nevertheless, it is necessary to verify the reliability of model predictions if designs based on these are also to be reliable. Here, it is considered that back-analysis is the basis of model calibration for reliable failure prediction. However, since neither the model (geology, constitutive behaviour, properties, pre-mining stress state) nor the true rock mass response can be determined with certainty, there is no objective way to calibrate a model. The reliability of the back-analysis itself is relative and dependent on the model and its parameters. Although back-analysis cannot guarantee unique solutions since different constitutive laws, numerical methods and boundary conditions may reach the same result; prediction reliability can be established by comparing results based on back-analysis of multiple predictions. Agreement in a few isolated cases is at best anecdotal. Reliability can only be established by using statistical techniques to compare the difference of many individual predictions with their average behaviour. Well-clustered results under a wide range of conditions would indicate reliable modelling predictions.

Since the reliability thus determined will depend on the model, the model itself can be modified (i.e. the geology, constitutive behaviour, properties, pre-mining stress state) to improve the clustering and minimize the sum total of the differences between individual predictions and their average behaviour. There is of course a limit to how good a fit that can ever be achieved owing to the inherent variability of the rock mass. As the variability in the outcome is a combination of the variabilities in the input, in general, the more parameters involved in prediction, the more variability should be anticipated in the outcome. This allows for direct comparison of the benefit of utilizing alternative models (i.e. elastic, plastic, creep, dynamic, etc.), with the cost of running them and the ever-increasing effort required to better characterise the geology, behavioural properties of the rock mass, and quantify the rock mass response.

The back-analysis approach described in this paper is restricted to situations where some sort of observable response occurs repeatedly. Situations where ground failure is routinely encountered (e.g. mining at high-extraction ratios or in weak ground) are ideal. This approach is therefore limited in its application to "green-field" sites where no calibration data is available, or in projects where the ground response is primarily elastic. While the applicability of the back-analysis method is limited, the statistical approach described in this paper is more generally applicable and can be used with the conventional empirical approach. The scope of this paper is limited to a discussion of the use of modelling for quantitative design. There are many other possible uses that are not covered here including parametric studies and sensitivity analysis.

2. Prediction using the conventional empirical method for rock mass strength estimation

Geological materials are often non-uniform, heterogeneous and anisotropic, as a result the stress, strength and other characteristics will vary from point to point. Although a mean value can be defined, there will be uncertainty as to the value that would be found at any given location. Repeated measurements demonstrate that the likelihood of finding a given value can be quantified in terms of probability. The range of variability is typically described by the coefficient of variation (C_v) for each parameter.

The reliability of a failure prediction can be determined using the standard methodology of probability and statistical analysis. To apply this, quantifying the mean and variability of both the rock mass strength (capacity function C) and stress predictions (demand function D) as shown in Fig. 1 is required.

Here the vertical axis represents the likelihood that various stress or strength levels will occur, and the width of the density distribution represents the variability. Prediction uncertainty is traditionally dealt with by applying a central factor of safety defined as

$$SF = \bar{C}/\bar{D} \tag{1}$$

where $\bar{C}$ and $\bar{D}$ represent, respectively, the mean capacity and demand.

Alternatively, these two distributions can be subtracted (i.e. the shaded area in Fig. 1) to provide a single function representing the overall probability of failure [3]. Although this later approach appears to be quite straight forward, application to design problems is not so simple. Direct determination of the probability distributions for stress predictions and in situ strength is essentially impossible. Reliable quantification of local variability would require detailed in situ sampling and measurements of induced

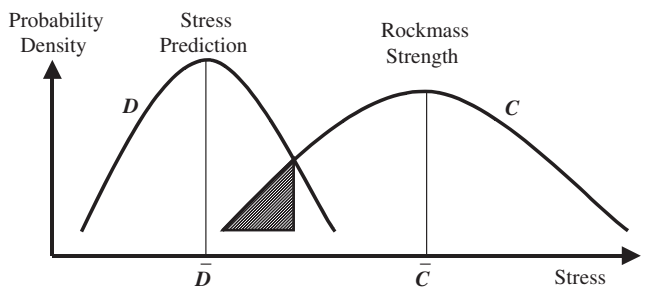

Fig. 1. Capacity versus demand.

stresses and rock mass strength. Owing to the large costs involved, this is an unrealistic objective particularly in mining environments. Hence, other means must be found to account for variability and to measure how representative or reliable a model prediction is.

In order to quantify the reliability of numerical modelling predictions, the variability of the capacity and demand functions needs to be estimated. The conventional empirical method for application of numerical modelling requires measurements of: (i) pre-mining in situ stress state; (ii) laboratory testing of rock samples; then, (iii) subsequent degradation of the intact strength to rock mass strength. The uncertainty in each of these parameter sets is combined with the uncertainty introduced by the numerical model to arrive at a coefficient of variation C_p for a prediction.

2.1. Estimating the capacity

The steps involved in estimating the capacity as shown in Fig. 2.

2.1.1. Laboratory strength

The variability found when making laboratory strength measurements can be expressed as a coefficient of variation. Values ranging from a low of 10% to a high of 40% are often reported for hard rocks [4]. Harr [3] tabulates coefficient of variation for a wide range of parameters and gives cohesive strength for soils at 40%. In the author's experience, values in the range of 20–30% can normally be achieved with good testing practices.

Although it is well known that the strength decreases markedly with increasing sample size [5], what is not known is whether the resulting coefficient of variation decreases as well.

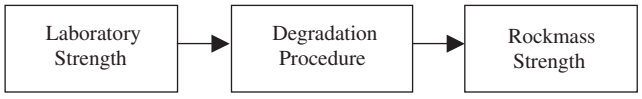

Fig. 2. Steps required to estimate the capacity.

2.1.2. Rock mass strength

Design techniques, including those involving numerical models, require an accurate representation of rock strength in the area or volume of interest. "Point" laboratory strength measurements must be degraded to approximate in situ rock mass strength magnitudes. This step can potentially introduce the greatest amount of uncertainty into our failure prediction since it is the relative magnitude of assumed pre-mining stress state to rock mass strength that will dictate whether failure is predicted or not.

There have been very few formal studies to quantify the level of uncertainty in this degradation process. One probable reason advanced for this is the fact that the actual rock mass strength is generally unknown. In situ testing at scales representative of the excavations is normally prohibitively expensive. Personal experience through back-analysis studies has shown that this degradation process is often unreliable and can result in significant under estimates of the rock mass strength.

Results from a recent paper by Villaescusa and Li [6] are reproduced in Table 1. Each column represents a different site. The rock mass unconfined compressive strength (*UCS*) has been estimated using various techniques that have been proposed to achieve degraded rock mass strengths. The magnitude of the coefficient of variation C_v for the degradation procedure can be estimated by calculating the standard deviation then dividing by the average for each column of predictions. This should provide some indication of the level of uncertainty inherent in this process. Two categories of C_v have been calculated for each column: C_{all} incorporates all 7 equations; $C_{9,10}$ excludes results from [9,10] (because as recommended by the above authors, these do not use the measured laboratory strength). The mean $\bar{C}_v$ has also been calculated for both categories.

The obvious lack of agreement between alternative methods for estimated field scale *UCS* suggests that the reliability of this procedure is low. Villaescusa and Li [6] further suggest that "Comparisons of the estimated parameters using rock mass classification and those used in numerical modelling (by mine site-based practitioners) … appear to estimate different rock mass compressive strengths for the same rock mass environment". Martin

Table 1
Estimated *UCS* (rock mass unconfined compressive strength in MPa)

GSI/RMR	75	50	30	75	65	24	$\bar{C}_v$
Hoek and Brown [7]	64.8	13	1.7	43	8.2	1	
Ramamurthy [8]	39.5	5.6	0.5	29	4.6	0.3	
Trueman [9]	45	10	3	45	24.7	2.1	
Singh [10]	57.4	22.7	10.8	57.4	39.6	8.7	
Kalamaras et al. [11]	57.4	18.8	2.4	42.1	9.7	1.2	
Sheorey [12]	43	6.6	0.6	31.5	5.2	0.3	
Hoek [13]	57	11.8	1.4	41.8	7.8	0.8	
C_{all}	18%	49%	123%	23%	92%	145%	75%
$C_{9,10}$	20%	48%	60%	18%	30%	57%	39%

Table 2
Estimated and measured *Coh* (rock mass cohesive strength in MPa) [4]

Site	CG1	CG2	FS1	M1	C_H	C_M	
GSI	74	65	65	54	60	46	
$Coh_{estimated}$	4.11	3.67	3.05	1.14	2.29	1.41	
$Coh_{measured}$	5.20	3.40	3.40	1.90	1.50	0.80	$\overline{Coh} = 2.70$
ΔCoh	−1.09	0.27	−0.35	−0.76	0.79	0.61	$s = 0.77$

and Maybee [14] find that the Hoek–Brown empirical method over estimates hard rock pillar strengths. Hoek [2] states that his own degradation procedure "works, more by good fortune than because of its inherent scientific merits". Furthermore, he states [15] "the user of the Hoek–Brown procedure or of any other equivalent procedure for estimating rock mass properties should not assume that the calculations produce unique reliable numbers".

In a study conducted by Cai et al. [4], rock mass strength is estimated using the Hoek–Brown method ($Coh_{estimated}$) and measured using in situ block shear tests ($Coh_{measured}$) at six different sites reported in Table 2.

The magnitude of the coefficient of variation for the degradation procedure can be estimated by calculating a standard deviation s from the differences between the estimated and measured values (ΔCoh), then dividing by a representative stress magnitude. Here we could use the average of the measured values $\overline{Coh}$ or the working stress level of 0–5 MPa suggested in the paper [4]. In either case, quite large values for C_v are indicated (i.e. 15–30%). The authors of this study conclude that the estimated cohesive strengths "are generally in good agreement with field data" despite the obvious large differences.

In a similar study by Kayabasi et al. [16], field scale values for the deformation modulus are estimated using several different techniques and also measured using a large number of plate loading tests. The reported root mean square errors for the estimated and measured values are reproduced in Table 3.

The magnitude of the coefficient of variation can be estimated by dividing the RMSE by the measured field scale deformation moduli (these were mostly measured in the range of 5–10 GPa), suggesting quite large values (i.e. well over 50%). The authors of this study conclude that the last three estimation techniques in the table "exhibited acceptable results. In fact these results are typical".

Degradation techniques inherently involve subjective assumptions to derive a reasonable estimate of in situ strength. Uncertainty occurs because of simplification of the actual complex mechanisms responsible for the degraded strength. Natural variability in the rock mass due to anisotropy, changing geology, pre-existing structure, etc. is normally averaged across the mine site in an attempt to derive a single strength envelope or at best, a series of strengths associated with broadly defined lithological units. The examples that have been summarized here demonstrate that existing empirical degradation techniques do not agree with one another, and also do not agree with field scale tests very well. Comments by the authors of the studies suggest that the results they obtained are generally expected, acceptable and typical.

Table 3
Root mean square errors (RMSE) for estimated modulus of deformability from the existing empirical equations [16]

Empirical equation	RMSE (GPa)	Number of tests
Bieniawski [17]	15.6	48
Asef et al. [18]	16.5	57
Serafim and Pereira [19]	8.9	9
Nicholson and Bieniawski [20]	5.6	57
Hoek and Brown [7]	8.9	33

From these results it would appear that the uncertainty associated with degrading laboratory test results to field scale is generally unknown. It may be appropriate to assign values in the range of 15–30% for the coefficient of variation for the degradation procedure. However, this number could be higher than this. It would appear that more work is required to confirm this.

2.2. Estimating the demand

The steps involved in estimating the demand are shown in Fig. 3.

2.2.1. Pre-mining stress state

Although the stress orientation and stress ratio can often be measured with reasonable confidence, it is notoriously difficult to determine the pre-mining in situ stress state magnitude by direct measurement. In addition, the local stress magnitudes can also vary quite widely across the mine site [21] due to a number of factors such as structures, and variable geology within a unit, amongst others.

Several hundred precision, temperature-controlled HI-cell (CSIRO hollow inclusion cell) measurements were made at Canada's URL (AECL Underground Research Laboratory). Analysis of these very high-quality measurements demonstrates that the coefficient of variation for stress magnitude is in the order of 20% [22]. It was found that additional measurements do not reduce this variability. At the same location, Martin et al. [23] calculate a coefficient of variation of approximately 20% for stress at depth, and indicate that this increases near ground surface.

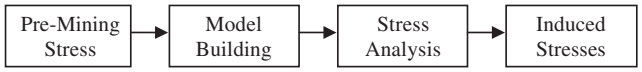

Fig. 3. Steps required to estimate the demand.

One could expect even higher variability at sites where the rock mass is not as uniform as at the URL.

Whether this uncertainty arises from natural variability of the rock mass or measurement error, this does represent the level of prediction uncertainty associated with the pre-mining stress state. The stress is normally averaged across the mine site in an attempt to derive a single stress state varying linearly with depth. Local deviations from the assumption of linearity are rarely taken into account in modelling, even though marked improvement in accuracy can be obtained [24]. This is primarily because of the lack of detailed information regarding the pre-mining stress distribution, but also because of the increase in complexity this introduces to the model.

2.2.2. Model building

Model geometries can today be built quite accurately owing to the high quality of modern surveying and the ease with which complex 3D geometries can be built with modern modelling packages such as Map3D [25]. With capacity to accommodate thousands of excavation units, there is no longer any reason for compromise in this regard. Excavations cause stress magnifications, thus accurate delineation of the geometry is necessary since pillar widths, proximity of excavations and 3D spatial location directly affect the stress redistributions that will be calculated during the analysis. Inaccurate representation of the geometry can provide a large contribution to uncertainty in the final stress prediction.

2.2.3. Stress analysis

Stress analysis calculates how the pre-mining stresses are modified due to mining. The relative magnitude of the pre-mining stress state with respect to the assumed rock mass strength directly controls the accuracy of the modelling predictions since failure predictions require a comparison of these to the strength. It is therefore necessary that the magnitude of both the pre-mining stress state and the rock mass strength be correctly specified to make accurate predictions.

There will also be uncertainty in this part of the process particularly if the rock mass is overstressed. Rock yielding will result in stress transfer that requires more complex inelastic models for correct simulation. This could be accommodated through incorporation of slipping faults or generalized yielding of the rock mass. Unfortunately the use of such models requires additional judgement and assumptions regarding a range of input parameters. In elastic stress models the only significant contributing factors are the geometry and pre-mining stress state; while in elasto-plastic models, the strength and flow rule become an integral part of the analysis. These assumptions change the way the model responds and directly influences the analysis results. Model results become loading path dependant.

It is unclear whether the increase in accuracy anticipated by use of a more complex material model is offset by the uncertainty introduced by the additional input parameters. It is conceivable that one could obtain less reliable predictions because of this (this is discussed in more detail in Section 5.6.2).

2.3. Combined effects—estimating C_p

It is timely to combine now the uncertainties associated with the laboratory strength, strength degradation to rock mass scale, pre-mining stress, and chosen modelling procedure into a single measure of the uncertainty in our predictive capability. This can be expressed as a coefficient of variation C_p. The coefficients of variation from the various sources can be easily combined by taking the square root of the sum of the squares of the individual standard deviations if it is assumed that normal distributions apply and that the contributions are uncorrelated [26].

Above it was shown that the uncertainty in laboratory strength is near 20%. Uncertainty in the degradation procedure may be in the range from 15% to 30% but could be higher than this. Combing these two figures gives an uncertainty in the rock mass strength estimate of 25–35%. Note that Hoek [13] calculates a value of 31% for the rock mass *UCS*. To determine C_p, the uncertainty in pre-mining stress (20%) and from the chosen modelling procedure must be added. Unfortunately the later contribution is generally unknown. Taking square root of the sum of the squares for all contributions it would appear that a coefficient of variation C_p of 30–40% should be expected. Realistic values are most likely higher than this, since some contributions have not been included and others are not known with any certainty.

The implication of various magnitudes of coefficient of variation on prediction accuracy is discussed in detail below.

3. Back-analysis approach

3.1. Background

To characterize the in situ variability of the stress and strength, an alternative to attempting to quantify the variability by actively conducting in situ measurements is to observe rock mass response during prior mining operations. This approach can also be used to assess the applicability of the chosen numerical modelling procedure (elastic, elasto-plastic, etc.) under similar conditions. This procedure could be considered to be an application of the *observational approach to design* [27].

Back-analysis of observations of rock mass response to advancing mining has the disadvantage that from an

experimental perspective, the conditions are uncontrolled. "It is futile ever to expect to have sufficient data to model rock masses in the conventional (for example in electrical or aerospace engineering) way" [1]. This is rather unsettling to practitioners who are used to conducting tests under well-controlled laboratory conditions. "Our confidence in the numerical models can be raised when they are successfully calibrated against well-controlled laboratory and in situ experiments" [28].

Under field conditions, precise measurements are often not possible or too expensive to be practical. However, in the author's opinion, this is more than offset by several major advantages of the observational approach. A wide range of realistic shapes and loading conditions are routinely exercised. A very large number of cases can be back analysed as part of the ongoing mining operation. Site specific values can be determined. Mining is of course true rock mass scale and uses the most representative shapes, stresses and conditions that are possible. Of course for this method to be viable, some sort of observable response must occur repeatedly. This can include (but is not limited to) a variety of stress-induced responses such as crack density, joint alteration, ground support requirements, blast-hole condition, ground stability, stand-up time, depth of over-break, dilution, micro-seismic activity, etc.

By using the same modelling tools for the back-analyses as will be used to make forward predictions, testing the applicability of the chosen numerical modelling procedure is occurring. In addition, use of this method bypasses the requirement for a strength degradation procedure. Back-analysis can be viewed as a procedure for quantifying the reliability of the entire predictive system rather than any of its individual components.

3.2. Quantification of uncertainty

To demonstrate the proposed methodology, consider the series of back-analyses of sill–pillar failures observed during mining operations. These are shown as solid diamond shapes in Fig. 4 [29–31].

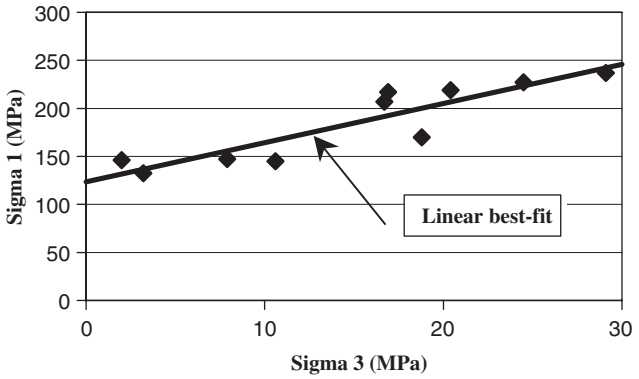

Fig. 4. Back-analysis of sill-pillar failures.

To obtain each of these points, a model was built and in this case analysed elastically using Map3D [25] to determine the stress state at the centre of each pillar at the observed time of failure. From these back-analyses results a best-fit strength envelope (shown as the solid-inclined line) using linear regression can easily be determined. It can be observed that the results are well clustered, as the difference between individual predictions and their average behaviour is small.

It is now time to consider how the coefficient of variation representing our predictive capability C_p from back-analysis results can be quantified. This is achieved by calculating the mean distance from the predictions to the best-fit line. For each back-analysis, the distance from any stress point to the best-fit line (excess stress) for a linear (Mohr–Coulomb) criterion is given by

$$\Delta\sigma_1 = \sigma_1 - UCS - q\sigma_3, \tag{2}$$

where σ_1 and σ_3 represent, respectively, the major and minor principal stresses, UCS and q represent, respectively, the rock mass unconfined compressive strength and slope of the best-fit line. $\Delta\sigma_1$ is positive above the line and negative below the line.

The standard deviation s can be written

$$s = \sqrt{\sum \Delta\sigma_1^2/(n-2)}, \tag{3}$$

where n represents the number of back-analysis points, and the summation is taken for all n data points. Here, $(n-2)$ represents the degrees of freedom used as a divisor to ensure an unbiased estimate. The best-fit line can be obtained by finding that values UCS and q that minimize the magnitude of s (i.e. linear regression) as shown in Table 4.

UCS and q can be related to the cohesion Coh and friction angle φ as follows:

$$UCS = 2\,Coh\tan(45+\varphi/2),$$
$$q = \tan^2(45+\varphi/2), \tag{4}$$

Table 4
Sill–pillar back-analysis

σ_1 (Mpa)	σ_3 (MPa)	$\Delta\sigma_1$ (MPa)
216.9	16.9	+24.42
170.0	18.8	−30.22
218.9	20.4	+12.16
236.7	29.1	−5.49
227.3	24.5	+3.85
206.9	16.7	+15.24
145.1	10.6	−21.71
132.6	3.2	−4.05
146.2	2.0	+14.44
147.1	7.9	−8.70
$\overline{\sigma_1} = 184.77$		$UCS = 123.61$
$\overline{\sigma_3} = 15.01$		$q = 4.075$
$s_1 = 40.29$		$Coh = 30.61$
$s_3 = 8.93$		$\varphi = 37.3^\circ$
$s = 18.39$		

s_1 and s_3 represent, respectively, the standard deviation of σ_1 and σ_3

$$s_1 = \sqrt{\sum(\sigma_1 - \overline{\sigma_1})^2/(n-1)},$$
$$s_3 = \sqrt{\sum(\sigma_3 - \overline{\sigma_3})^2/(n-1)}, \quad (5)$$

where $\overline{\sigma_1}$ and $\overline{\sigma_3}$ represent the mean values. These are conveniently related to the standard deviation s (Eq. (3)) by

$$s = \sqrt{(s_1^2 - q^2 s_3^2)(n-1)/(n-2)}. \quad (6)$$

For the back-analysis results shown in Fig. 4 and Table 4, a standard deviation 18.39 MPa can be calculated. It is usually more meaningful to express this as a coefficient of variation by dividing by a representative stress magnitude. Here, the mean value of σ_1 (185 MPa) will be used giving

$$C_p = s/\overline{\sigma_1} = \pm 10\%. \quad (7)$$

s and hence C_p are site-specific characteristics determined from back-analysis that quantify how well the entire modelling procedure is performing. The contributions from the variability of the pre-mining stress and rock mass strength, material heterogeneity and also errors introduced by the modelling procedure are inherently included. This later contribution has not been quantified before. These parameters represent the uncertainty one has in predictive capability. Note that C_p is the same parameter whose evaluation was attempted in Section 2.3. Recall that when using the conventional empirical approach we anticipated that a value in excess of 30% would be appropriate. The surprisingly low value of 10% obtained here demonstrates why the back-analysis approach is so appealing.

3.3. Error table minimization procedure applicable to inelastic models

The above discussion has been presented from the perspective of quantifying the uncertainty in elastic modelling results. This methodology can be applied to the more general case of minimizing prediction errors with respect to stress ratio or orientation, and to inelastic modelling with a few minor adjustments.

In order to proceed, the standard error requires redefining. Consider a series of back-analyses conducted using an inelastic model with m different strength parameters (e.g. different values for the peak and residual UCS and q). The excess stress given as Eq. (2) can be used when the stress prediction lies below the strength envelope (large dot on the left-hand side of Fig. 5). In cases where yielding occurs (large dot on the right-hand side of Fig. 5), we can use the extrapolated elastic over-stressing as illustrated in the figure.

For each of the n back-analysis locations, an error table of entries for $\Delta\sigma_1$ as a function of each value of UCS and q must be calculated. Next, the square root of the sum of the squares divided by n–m, for the entries from each table is taken. This defines the error distribution s as a function of UCS and q analogous to Eq. (3) as shown in Fig. 6. Since each back-analysis represents a stress level exactly when failure was observed to occur, the best fit is located where the total error s is minimized. This could readily be determined by interpolation within the total error table, corresponding to the regression procedure used above.

To avoid extrapolation, at least 2 values for each of UCS and q (i.e. 4 separate analyses per back-analysis point) must be considered. It is preferable to include at least 2 points below and 2 points above the best-fit strength envelope. For an inelastic analysis, it is most likely that the error surface will not be highly non-planar, possibly

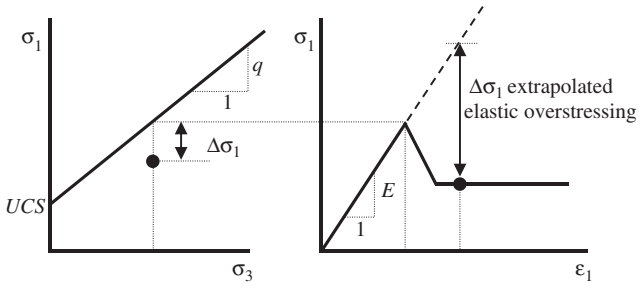

Fig. 5. Standard error definition for an inelastic model.

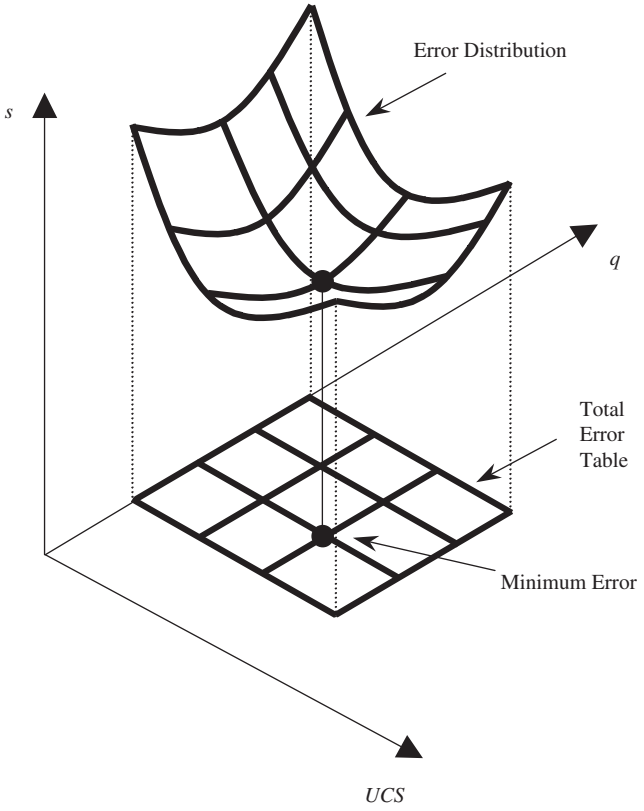

Fig. 6. Error table distribution.

discontinuous and loading path dependant, therefore many more points may be required to identify the best fit accurately. For example, considering 4 values for each of value of *UCS* and q would require 4 to the power 2, or 16 separate analyses per back-analysis point. With 10 back-analysis points, this requires 160 inelastic stress analyses. Once the best-fit values for *UCS* and q are determined, the statistics given in Eq. (5), (6), and (7) are readily calculated.

If a more complex strain softening material model is considered then there will be additional parameters that must be minimized. These can include: post-peak strength (2 parameters), dilation rate (1 parameter), strain-softening rate (1 parameter), pre-peak moduli (2 parameters), post-peak moduli (2 parameters) and more. Sensitivity to loading path must also be considered. For each additional parameter, a dimension to the error table must be added. For example with 6 parameters (i.e. m equals 6), generation of a 6 dimensional error table is needed. Requiring 4 evaluations for each parameter, this would require 4 to the power 6, or 2096 separate analyses per back-analysis point. With 10 back-analysis points, this requires 20,960 inelastic stress analyses. Realistically one may make-do with fewer back-analyses, however it is evident that error minimization will require many analyses.

This same approach could be used to minimize prediction errors with respect to stress ratio or orientation (or any other parameter) by specifying these later parameters as additional dimensions in the error table. This could be done for either elastic or inelastic modelling. The approach provides an opportunity to make use of detailed observations of in situ response whether visual or from instrumentation. This methodology can be considered to be a generalization of direct error minimization techniques such as under-excavation [32,33].

Even without investing any effort in error minimization, the statistics given in Eqs. (5), (6) and (7) can still be readily calculated for any specific value of *UCS* and q. In this later case only one analysis for each back-analysis point would need to be conducted. Although in such a case the best-fit strength parameters would not be used, predictions could still be made with these statistics.

3.4. Probability of failure

If the assumption is that σ_1 is normally distributed, the probability of failure P_f can be readily calculated by integrating the overlapping area shown in Fig. 1

$$P_f = N(\Delta\sigma_1/s), \tag{8}$$

where N is a function that represents the area under the standardized normal curve (Table 5 with $n = \infty$).

For small data sets the calculated values for the best-fit strength envelope and standard deviation are only uncertain estimates. This extra uncertainty can be incorporated as a function of sample size n by using the t distribution [26]

$$P_f = T\left(\frac{\Delta\sigma_1}{sg}, n-2\right), \tag{9}$$

where T is a function that represents the area under the standardized t distribution curve for $n-2$ degrees of freedom and g is given by

$$g = \sqrt{1 + \frac{1}{n} + \frac{(\sigma_3 - \overline{\sigma_3})^2}{(n-1)s_3^2}}. \tag{10}$$

This later factor quantifies the additional uncertainty associated with being at the extremities of the data range in terms of σ_3. It represents an increasing lack of knowledge when σ_3 deviates a long way from the mean and thus prevents invalid interpolations.

Given P_f, the confidence interval can be determined by use of the inverse of this function

$$\Delta\sigma_1 = sgT^{-1}(P_f, n-2). \tag{11}$$

The 1%, 5%, 95% and 99% confidence intervals determined using Eq. (11) are plotted for reference in Fig. 7. Here the solid line represents the best fit (P_f of 50%) and the dashed lines represent the various confidence intervals.

The results for the t distribution are very similar to the normal distribution except there is increased uncertainty (i.e. the confidence intervals are further from the mean) near the limits of the σ_3 range and for small values of n as shown in Table 5.

Table 5
t distribution [26]

P_f (%)	$n = 3$	$n = 5$	$n = 10$	$n = 20$	$n = \infty$
0.1	−318	−10.2	−4.50	−3.61	−3.09
1	−31.8	−4.54	−2.90	−2.55	−2.33
2.5	−12.7	−3.18	−2.31	−2.10	−1.96
5	−6.31	−2.35	−1.86	−1.73	−1.65
10	−3.08	−1.64	−1.40	−1.33	−1.28
50	0	0	0	0	0
90	+3.08	+1.64	+1.40	+1.33	+1.28
95	+6.31	+2.35	+1.86	+1.73	+1.65
97.5	+12.7	+3.18	+2.31	+2.10	+1.96
99	+31.8	+4.54	+2.90	+2.55	+2.33
99.9	+318	+10.2	+4.50	+3.61	+3.09

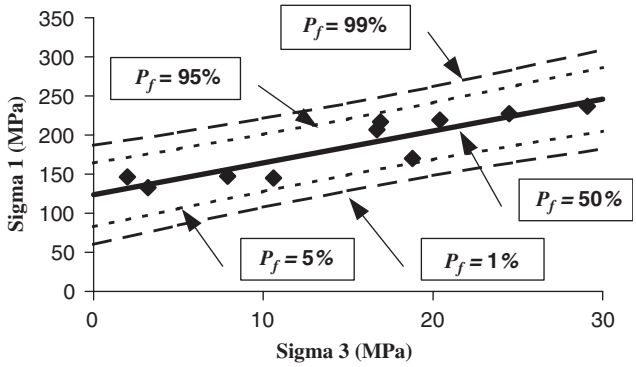

Fig. 7. Confidence intervals defined in Eq. (11).

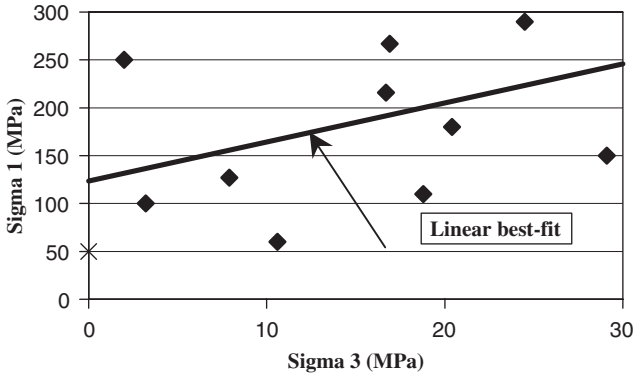

Fig. 8. Back-analysis of sill-pillar failures—fictitious data ($C_v = 40\%$).

Note that for large values of n, the t distribution asymptotically approaches a normal distribution and g asymptotically approaches unity, hence the probability of failure is given by Eq. (8).

3.5. Safety factor

The definition for safety factor from Eq. (1) can be written

$$SF = (UCS + q\sigma_3)/\sigma_1. \tag{12}$$

In Fig. 7 it is apparent that certain levels of stress can be directly associated with probability of failure. This can be expressed by substituting for Eq. (2)

$$SF = (UCS + q\sigma_3)/(UCS + q\sigma_3 + \Delta\sigma_1). \tag{13}$$

An approximate relation can be determined between safety factor and probability of failure by evaluating Eq. (13) at the mean $(\overline{\sigma_1}, \overline{\sigma_3},)$ and rearranging to obtain

$$\overline{SF} = 1/(1 + \Delta\sigma_1/\overline{\sigma_1}). \tag{14}$$

Now substituting Eq. (7), (10) and (11) the following can be derived

$$\overline{SF} = 1/\left[1 + C_p T^{-1}(P_f, n-2)\sqrt{1 + 1/n}\right]. \tag{15}$$

For large values of n this simplifies to

$$\overline{SF} = 1/[1 + C_p N^{-1}(P_f)]. \tag{16}$$

These results demonstrate that safety factor, uncertainty (i.e. C_p), and probability of failure are closely linked. Safety factor determined in this way can be used as a simplified approximation to the more rigorous calculation of probability of failure described in the previous section.

3.6. Back-analysis approach

From the back-analysis results illustrated in Fig. 7, with enough data, it is possible to evaluate whether or not the failure envelope should be a straight line, curved or otherwise. The goodness of fit can be readily assessed. From this simple procedure it is immediately obvious whether the model is working or not. There is one compelling advantage to approaching failure prediction in this manner. If the magnitude specified for the pre-mining stress state was too high or too low, the back-analysis results (and hence the best-fit line) would compensate simply by shifting up or down. This relieves the burden of attempting to accurately determine the absolute magnitude of stress and strength through direct measurement.

The back-analysis results plotted with large scatter as shown in Fig. 8 (coefficient of variation of 40%), it would indicate that the model is not working in this situation. This can result from technical modelling problems such as 2D versus 3D, non-convergence, instability, chaotic behaviour, geometric construction errors or numerical approximation errors. Perhaps, incorporation of important geological features such as changing lithology is required. It may be that the pre-mining stress state orientation or stress ratio assumption is incorrect. It is also possible that an elasto-plastic, yielding model with slipping faults or yielding pillars is required. Another possibility is simply that the back-analysis results have been collected from different lithological units and been superimposed on the same σ_1 versus σ_3 plot. Creating a separate plot for each lithological unit may be all that is required to resolve the different strength clusters. In this case, we must decide if the extra costs involved with refining our procedure are worth the benefits of more accurate predictions.

In any case, if results such as those pictured in Fig. 8 were obtained, it is obvious that there would be little basis for prediction. For this reason, it is necessary to provide variability information with all modelling predictions. It is important to note that the fundamental difference between Figs. 7 and 8 is only the variability in the results. In both figures, the same solid line represents the best fit to the results obtained by linear regression. This extreme has been presented to demonstrate the necessity of presenting variability information along with the mean values.

4. Practical applications of variability concepts

With the failure envelope and the statistics representing predictive uncertainty determined, it is now timely to apply this to a few examples to determine what the limits to the accuracy of failure prediction are.

4.1. Broken ground depth

This methodology can now be applied to ground support design. The dead weight that the support needs to suspend is the ground that has undergone stress driven failure, and hence, should correspond to the ground that has been stressed beyond the rock mass strength [34]. Consider the vertical section taken across the back of the 8 m wide slot shown in Fig. 9. This figure illustartes only a small part of the mining excavations. Contours of probability of failure using Eq. (9) are readily plotted as shown in Fig. 10 using the information determined in Table 4. In Fig. 11 it is shown how the depth of failure varies along the length of the same highly stressed back. Note the increased uncertainty indicated by the thicker band of contours as you approach the stope face.

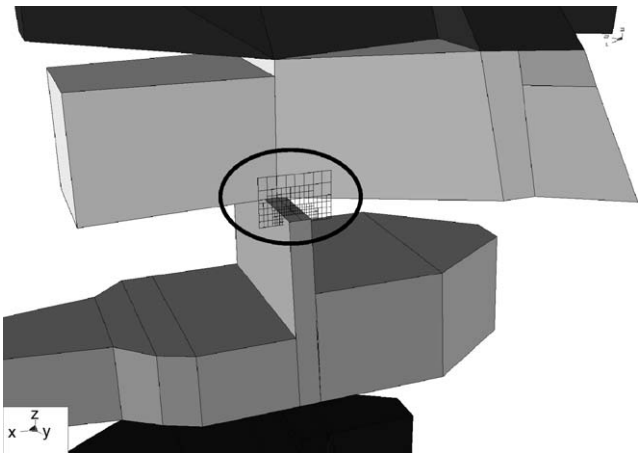

Fig. 9. Highly stressed back.

For the cross-section shown in Fig. 10, a prediction can be made with 90% confidence (95% minus 5%) that the depth of failure is between 1.54 and 3.0 m. Other confidence intervals can easily be scaled off the contours if desired. Note that the position of these contours is directly dependant on the magnitude of s used in Eq. (9). For a smaller value of s, a narrow range of uncertainty would be found. Recall also that the value of s is a site-specific characteristic determined from back-analysis that quantifies how well the modelling procedure is functioning.

Details of the stresses in Fig. 10 are shown as solid diamonds in Fig. 12, 13 and Table 6.

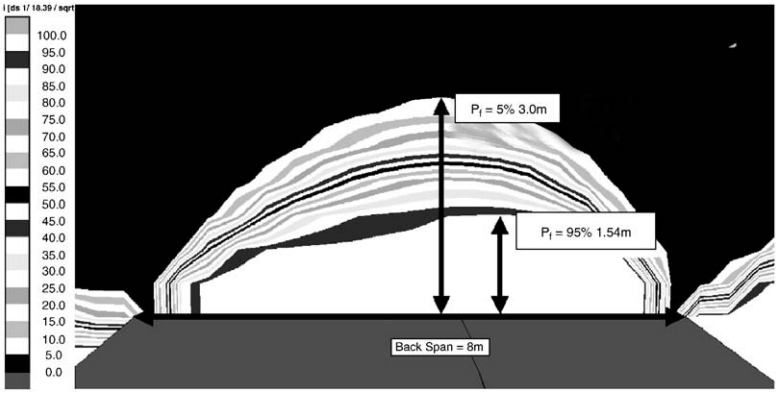

Fig. 10. Contours of P_f—cross-section through the highly stressed back.

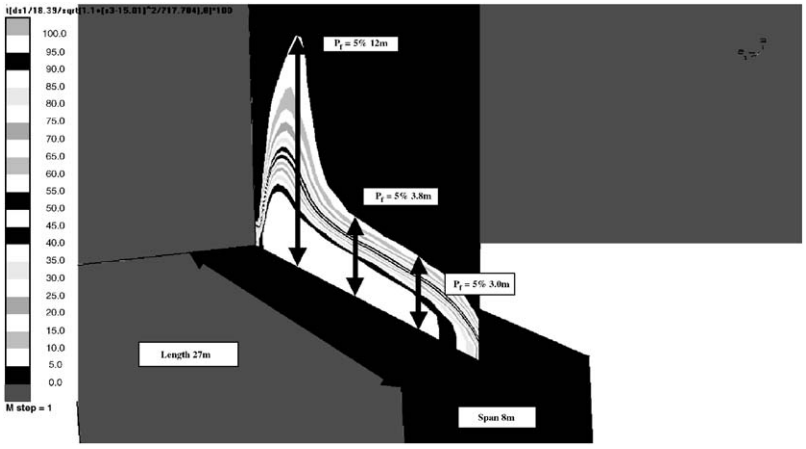

Fig. 11. Contours of P_f—long-section along the highly stressed back.

Results in Fig. 13 confirm the predicted depth of failure between 1.54 m and 3.0 m made from the contours in Fig. 10. This figure also illustrates that the results are well represented by a normal distribution (shown as the hollow circles) with a mean of 2.23 m and standard deviation of 0.442 m.

4.2. Crown-pillar failure

The results provided in Fig. 7 and Table 4 show back-analyses of failures from the silling out stage at three different levels (over 2 km depth) of Inco's Creighton Mine during the mid-1980s (for example see Fig. 14). In all cases, the failures were obvious as the pillars failed by bursting. These bursts resulted in the displacement of considerable quantities of material varying from 7 to 200 ton.

The initial silling created a 59 m wide (in the vertical direction), horizontally oriented crown-pillar as shown in Fig. 15. After several years, the mechanized cut and fill mining had progressed to create a narrowing crown-pillar that eventually failed violently at 35 m width and approximately a 2 to 1 height to width ratio. Let us now consider how the above reliability concepts can be applied to predict the onset of crown-pillar failure.

Fig. 16 shows the stress state predicted from elastic modelling for various crown-pillar widths. The solid diamonds correspond to 8 m intervals, representing two cuts each. Details are given in Table 7 and presented in Fig. 17.

Fig. 17 shows that a prediction can be made with 90% confidence (95% minus 5%) that the crown-pillar will fail

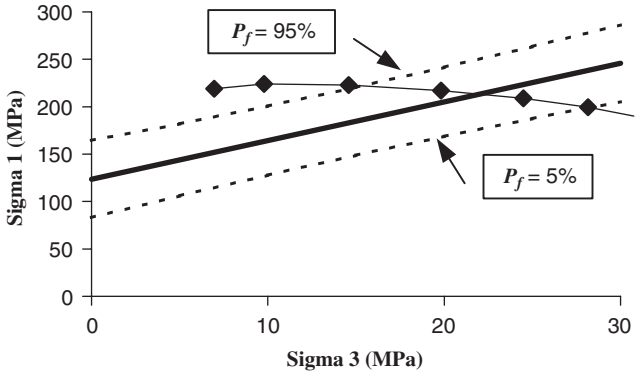

Fig. 12. Stresses above the highly stressed back at 0.5 m intervals.

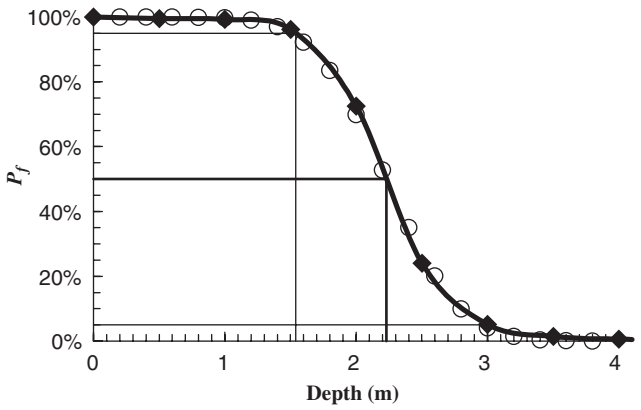

Fig. 13. Probability of failure above the highly stressed back.

Table 6
Details of stresses in the highly stressed back

Depth (m)	σ_1 (MPa)	σ_3 (MPa)	$\Delta\sigma_1/(sg)$	P_f
0.5	218.8	6.96	+3.31	99.5%
1	224.1	9.79	+3.07	99.2%
1.5	222.8	14.6	+2.04	96.2%
2	217.0	19.8	+0.62	72.5%
2.5	208.7	24.5	−0.74	24.0%
3	199.5	28.2	−1.85	5.12%
3.5	190.1	30.8	−2.68	1.40%
4	181.0	32.3	−3.30	0.54%
5	164.7	33.1	−4.10	0.17%
6	151.2	32.2	−4.60	0.09%

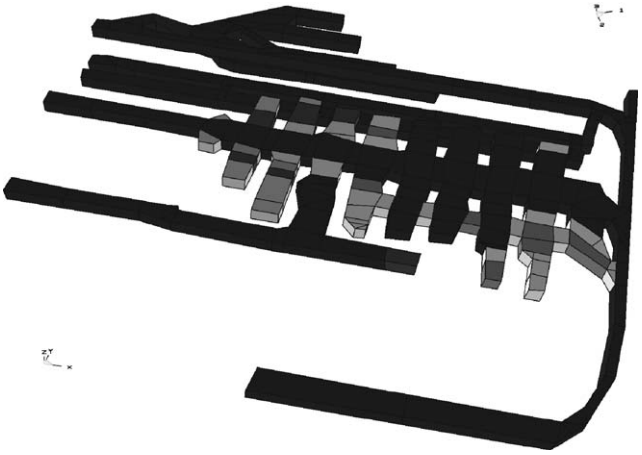

Fig. 14. Geometry used for the sill-pillar back-analysis.

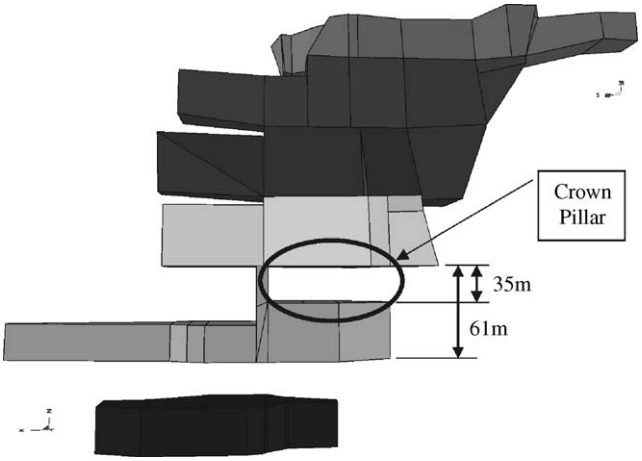

Fig. 15. Location of the crown-pillar failure.

when its width is between 8 and 54 m, or ±23 m. Other confidence intervals can be determined from Fig. 17 if desired. This figure also illustrates that although the results are skewed, they can be approximated by a normal distribution (shown as the hollow circles) with a mean of 38.3 m and standard deviation of 13.8 m.

Note that the positions of the confidence intervals in Fig 16 are directly dependent on the magnitude of s used in Eq. (9). For a smaller value of s, a narrow range of uncertainty would be found. Recall also that the value of s is a site-specific characteristic determined from back-analysis that quantifies how well the modelling procedure is functioning. These results are discussed in more detail below.

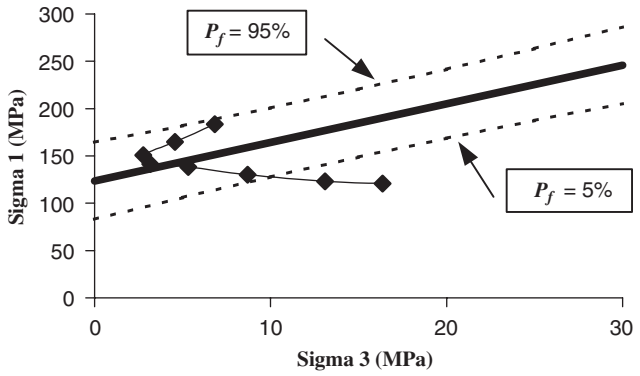

Fig. 16. Stress in the crown-pillar for various pillar widths.

Table 7
Details of stresses in the crown–pillar

Width (m)	σ_1 (MPa)	σ_3 (MPa)	$\Delta\sigma_1/(sg)$	P_f
66	120.7	16.4	−3.57	0.36%
59	123.1	13.1	−2.78	1.20%
51	130.6	8.69	−1.46	9.10%
43	138.2	5.33	−0.36	36.6%
35*	141.3	3.17	+0.22	58.3%
27	150.6	2.76	+0.74	75.8%
19	164.8	4.56	+1.09	84.6%
11	183.4	6.82	+1.59	92.5%

*Failure.

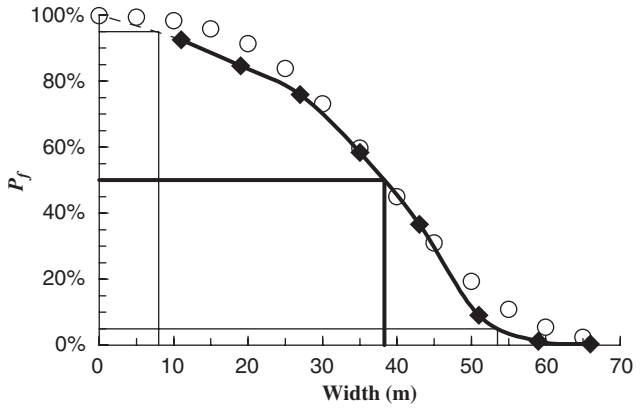

Fig. 17. Probability of failure in the crown-pillar.

5. Discussion

This paper has demonstrated how to quantify the accuracy of failure predictions using numerical modelling. By considering the uncertainty of the various contributing factors, many interesting and useful results have emerged.

5.1. Prediction reliability

The probability of failure that is acceptable is of course dictated by the project. Harr [3] notes that most civil engineering systems are designed with a probability of failure between 1% and 5% (i.e. a reliability of 95–99%). Daehnke et al. [35] recommend using a 95% confidence level (5% probability of failure) for the South African gold mining industry. In non-entry mining operations it may be possible to sustain higher probabilities of failure owing to short-term stability requirements, or where failure may simply mean rehabilitation rather than catastrophe.

In the development of the methodology above, the discussion thus far has been limited to the application of normal distributions. This simplicity has lead to many intuitive insights. If there was a need to extend this to non-normal probability distributions, Rosenbleuth's [36] point estimate method, or Monte Carlo methods could be adopted. This has already been demonstrated in much detail [15,23].

5.2. Safety factor

Two problems can be identified with the definition of safety factor. The same magnitude of safety factor corresponds to different probabilities of failure depending on the magnitude of C_p. For example, consider two different sites with two different sets of back-analysis results. Using Eq. (10), (11) and (13), we can determine that at σ_3 equals zero, a P_f of 5% requires a SF of 1.49 (with C_p equal to 10% and n equal to 10). If at a different site (with a different set of back-analysis results) we determined that C_p was equal to 20%, the SF required would be 2.92. Widely different values for safety factor are required to give the same probability of failure depending on the local site specific value of C_p.

An additional problem is that the value of safety factor required also depends on the magnitude of σ_3. For example, as above we can determine that at σ_3 equals zero, a P_f of 5% requires a SF of 1.49 (with C_p equal to 10% and n equal to 10). With the same C_p, but at σ_3 equals 15 MPa, the SF required would be 1.24. This effect is too large to be ignored.

These problems are a direct result of the definition of safety factor and imply that the same numeric value of safety factor can represent different probabilities of failure and hence different levels of safety: definitely an oxymoron.

This safety factor only provides a unique measure of safety under very restrictive conditions. For these reasons it is recommended that probability of failure be used.

5.3. Conventional empirical method for rock mass strength estimation

The conventional numerical modelling approach requires measurement of the laboratory strength and subsequent degradation to rock mass scale. The uncertainty in these later parameters is combined to determine an estimate of the coefficient of variation for our rock mass failure criterion. To this, the uncertainty associated with the pre-mining in situ stress state and the stress analysis procedure are combined to arrive at the coefficient of variation C_p for our predictive capability.

Taking into account these various contributions, a final C_p of 30% or more should be used. Using Eq. (16) we can determine that this corresponds to a $\overline{SF}$ of 1.6, 2 and 3.3, respectively, for 10%, 5% and 1% probability of failure. This estimate for C_p would appear to be realistic, as the resulting safety factor does not appear to be out of line with accepted practice. Safety factors of 2–2.5 are common in building design. Obert and Duval [37] recommend using values from 2–4 for mine pillars and sidewalls with a relatively short lifetime, and 4–8 for openings with a long lifetime. Lower values are generally recommended where frictional effects dominate. For rock slopes Hoek and Bray [38] quote 1.5 for cohesive strengths and 1.2 for frictional strength. Values from 2 to 4 are required for gravity dams, 4 for concrete arch dams, and 1.2–1.5 for embankment dams [39].

This large uncertainty does not allow for very well optimised designs.

5.4. Back-analysis approach

An alternative methodology is proposed based on the *observational approach to design*. Back-analysis results are used to determine a best-fit failure envelope that characterizes the combined uncertainty in pre-mining stress, rock mass strength and applicability of the chosen modelling technique (elastic, elasto-plastic, etc.). The compelling advantages to this approach are many:

(1) The ratio of pre-mining stress to rock mass strength is inherently determined as part of the calibration process. This relaxes the sensitivity of predictions to assumptions regarding the pre-mining stress state and removes the uncertainty involved with the procedure for degradation of laboratory strength to field scale values.
(2) The uncertainty introduced by the chosen modelling procedure is automatically included. Well-clustered back-analysis results indicate when reliable modelling predictions are obtained. Highly scattered results indicate problems with one or more assumptions.
(3) The coefficient of variation for the predictive capability is readily determined and can result in values for C_p as low as 10%. This corresponds to a $\overline{SF}$ of 1.24 (for 5% probability of failure with n equal to 10). This is considerably lower than what is obtainable from the conventional empirical approach (Section 5.3), and is a site-specific value that needs to be determined through back-analysis. This improved reliability allows for better-optimised designs.
(4) Back-analysis provides an opportunity to refine input parameters by seeking values that reduce the scatter of the clustering about the best fit. Using the error table minimization procedure, this can be applied to both strength parameters and far field stress state parameters. With enough data, it should be possible to evaluate whether or not the failure envelope should be a straight line or curved. Local variations of the strength and pre-mining stress state could be characterized.
(5) This can be applied to any situation where some sort of observable, stress-induced response occurs repeatedly. Situations where ground failure is routinely encountered (e.g. mining at high extraction ratios or in weak ground) are ideal. Envelopes corresponding to a variety of responses such as crack density, joint alteration, ground support requirements, blast-hole condition, ground stability, stand-up time, depth of over-break, dilution, micro-seismic activity, etc., can all be developed.

5.5. Conclusive failure predictions

An important question here is whether the moment of failure can be predicted. In the earlier crown-pillar back-analysis, it was determined that the crown-pillar will fail when its width is in the range of 8–54 m or ± 23 m (90% confidence, C_p of 10%, n equal to 10). This uncertainty represents well over half the mining life of the pillar. In terms of predicting the exact moment of failure, this is clearly not a very useful prediction. It will be shown below (Fig. 19) that with a C_p greater than 20% we would not even be able to predict with any certainty, whether the pillar would fail.

In spite of our inability to predict the moment of failure, we are quite certain that the crown-pillar is going to fail in the middle 11 cuts. A more in-depth analysis [30,31] demonstrates that not only can we be confident in predicting this failure, but by use of energy release rate calculations, it can also be shown that the failure will be a violent rock burst.

5.6. Refining accuracy of predictions

The accuracy of predictions can be refined in two ways, increase n and reduce C_p. If a larger number of back-analysis results were available (i.e. larger n), the results in Table 7 and Fig. 17 could be recalculated. From this it could be predicted with 90% confidence that the crown-pillar will fail when its width is between 12 and 52 m or

± 20 m. This is not significantly different from the prediction made with n equal 10 (i.e. ± 23 m). Close examination of Table 5 and Eq. (10) shows that the effect of sample size n is only dominant for very small samples or very large confidence intervals. To demonstrate this, the width range has been calculated for a number of different values for n as illustrated in Fig. 18.

There is very little to gain by back analysing more than 10 data points in this case. If higher levels of confidence were required, sensitivity to the magnitude of n would be more pronounced.

Had the back-analysis results been more scattered giving for example a C_p of 20%, the results in Table 7 and Fig. 17 could be recalculated. From this it could be determined that the uncertainty would encompass the entire pillar, as shown in Fig. 19. This means that there is uncertainty as to whether the pillar is going to fail or not.

From this figure it can be observed that uncertainty as to when the crown-pillar will fail is directly proportional to the variability in our back-analysis results C_p (recall that in Section 2.3 we estimated a C_p of 30% or more for the conventional empirical approach).

From Fig. 19 it can be determined that in order to double the prediction accuracy (i.e. from ± 21 to ± 10.5 m), a C_p of 4.2% would be needed. A further reduction in uncertainty to ± 5 m would require a C_p of 1.5%. It seems unlikely that back-analysis results that matched observations to the level of detail necessary to result in such a low value are obtainable. This implies that high-accuracy conclusive predictions are not possible in this case.

5.6.1. Reducing C_p

For a given site, there is an inherent background level of uncertainty due to the variability associated with the in situ stress, strength and changing geology. The magnitude of this contribution could be reduced if we could spatially correlate our rock mass failure criterion to match these changes.

Here, consideration could be given to incorporating non-homogeneous geological details into the model, aiming for a better match with actual changing conditions across the site. To achieve this it may be as simple as introducing zones with different stiffness or pre-mining stress states. Conversely, it could be as complicated as fault slip simulation to model important structural features.

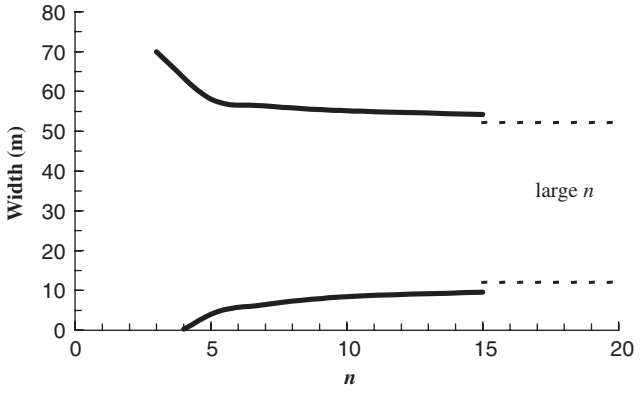

Fig. 18. Predicted crown-pillar failure width-range for various sample sizes (90% confidence interval).

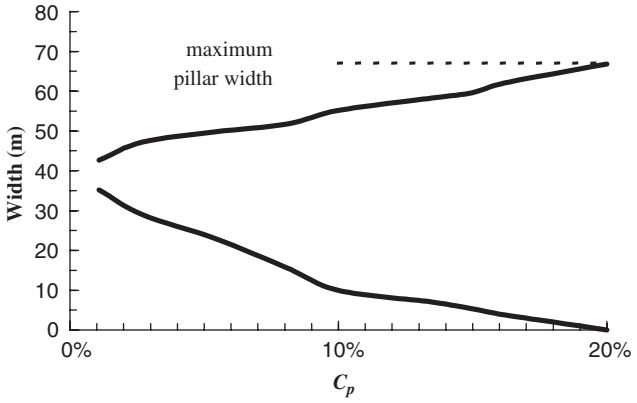

Fig. 19. Predicted crown-pillar failure width-range for various values of C_p (90% confidence interval).

A simpler alternative to modelling the geological complexity is to determine a heterogenous rock mass strength distribution. Consistent success has been achieved [40] by broadly defining lithological units across a mine site. This benefit may arise due to actual varying rock mass strengths, or alternatively could reflect real but unknown variations in the magnitude of the pre-mining stress state. This later explanation is quite likely if one keeps in mind that numerical models actually calculate stress magnifications of the pre-mining stress state rather than absolute values. In a study by the author [24], it was found that when local deviations from assumed linear variation of the pre-mining stress magnitude with depth were measured and included in the model, remarkably accurate pillar stress predictions were obtained.

Recently, attempts have been made to reduce C_p by directly modifying the simulated rock mass response by physically loading the numerical model though incorporation of seismicity [41,42]. This technique imposes the inelastic deformations implied by observed seismicity directly into the model. So far these have achieved limited success.

5.6.2. Numerical modelling technique

When large values of C_p are found from back-analysis, it is possible that this arises from use of an inappropriate modelling technique. Obvious causes can include ill-posed models (2D versus 3D, non-converged, unstable, chaotic systems), geometric construction errors or numerical approximation errors (inadequate discretization). These technical problems while important, are not at issue here.

In heavily loaded mines where significant stress transfer occurs as a result of yielding ground, elastic models may not provide accurate predictions. The rudiments of stress transfer could be incorporated into an elastic model simply

by excavating failed pillars. Introducing either inelastic fault slip or elasto-plastic yielding to affect stress transfer could be considered. Block models could be used if ravelling is a dominant feature. However, caution is required if proceeding on this course. In addition to adding more complex simulation capability, more assumptions are added which can result in an increase in uncertainty.

With more parameters to adjust, it becomes progressively easier to achieve any specific desired result. However, this does not guarantee unique solutions since with more parameters, many combinations of parameters may reach the same result. Prediction reliability can only be established by comparing results based on back-analysis of multiple predictions under a wide range of conditions. Reliability can then be established using statistical techniques by comparing the difference of many individual predictions with their average behaviour. Well-clustered results under a wide range of conditions would indicate reliable modelling predictions. Despite good intentions, it is entirely possible that less reliable predictions can be the outcome. This often happens because the costs are simply too large to complete the number of back-analyses required to properly calibrate the large number of parameters used with inelastic models. Whereas a few dozen analyses may suffice to test for good clustering and characterize the uncertainty in an elastic model, many thousands of analyses (Section 3.3) can be required for inelastic models.

Even though a complex model has the potential to provide more accurate predictions, given a limited budget and time, one must often choose between a well calibrated simple model or a poorly calibrated complex one. More information may be gained by running many simple models rather than a few complex ones. The importance of this issue is discussed by Starfield and Cundall [1] whose paper focuses entirely on this issue.

If geological variability is primarily responsible for the mismatch between back-analysis results and observed in situ response, then more complex inelastic modelling will not improve prediction accuracy. Even though Creighton mine is one of the most highly stressed mines in the world (pre-mining stresses exceed 100 MPa at depth), it is doubtful that inelastic modelling would significantly reduce the C_p of 10% determined from elastic modelling for the back-analysis results presented in Fig. 7. The well-clustered nature of these results indicates very reliable modelling predictions. Engineering effort would probably be better directed in attempting to spatially correlate our rock mass failure criterion to match the heterogenous rock mass strength distribution as discussed in Section 5.6.1 above.

5.7. Verification of conventional empirical strength estimates

Once a rock mass strength estimate based on the conventional empirical method has been obtained, a comparison between predictions and observed ground response can be made. The question here is how to proceed if in doing this it was found that the strength estimate was inaccurate and results as shown in Fig. 20 were obtained.

The results shown in this figure appear to demonstrate a possible systematic error. It is important to note that the modelling procedure is producing well-clustered consistent results with a relatively small amount of variability (these are the same back-analysis results considered above in Figs. 4 and 7 that give a C_p of only 10%), yet it is clear that predictions are unable to be made with any accuracy. An error should be declared in one or more of the modelling parameters.

Although there are many possible sources of error, the possibility of rock mass heterogeneity and inelastic effects can be discounted due to well-clustered nature of the predictions. The most likely cause would be that either the pre-mining stress magnitude or that the strength degradation procedure is in error. It would be necessary at this point to either revise the strength estimate upwards, or decrease the pre-mining stress magnitude until a match of the back-analysis results could be made.

5.8. Recommended procedure for making predictions

In view of the above discussion, recommendations regarding application of modelling to design problems can be made:

(1) *Accurate geometry*. A good representation of the geometry is necessary in order that stress redistributions are accurately calculated during analysis. In any but the simplest of problems this will require three-dimensional geometric representation since the proximity of excavations and 3D spatial location directly affect the calculated stress concentrations.
(2) *Simple model*. Initially, the simplest possible modelling approach should be adopted so that the least number of parameters need to be estimated. Homogeneous, elastic modelling is the best option since the only significant parameter that must be specified is the far-field stress state. It is really only the stress orientation and ratio

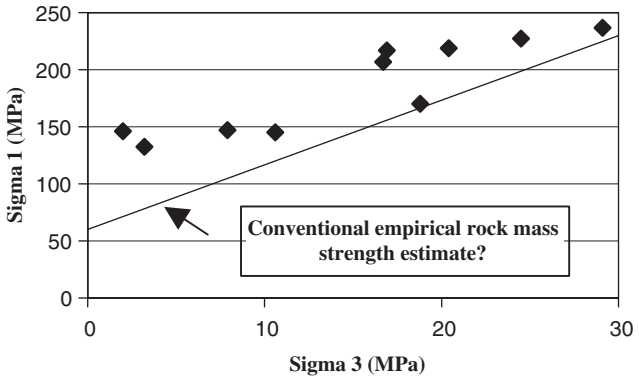

Fig. 20. Conventional empirical rock mass strength estimate compared to observed failures.

that is important here, since the magnitude with respect to the rock mass strength is best calibrated through back-analysis.

(3) *Back-analysis—C_p*. Back-analysis of observed in situ response should be conducted to determine a representative rock mass strength envelope and corresponding coefficient of variation C_p. This need not be restricted only to stress driven failure events such as pillar failures. By back-analysing non-events (i.e. cases where no failure is observed), a lower limit to the strength can be quickly established. Envelopes corresponding to a variety of stress-induced responses such as crack density, joint alteration, ground support requirements, blast-hole condition, ground stability, stand-up time, depth of over-break, dilution, microseismic activity, etc., can all be considered.

(4) *Predictions—P_f*. Predictions should be made with specified reliability using the probability-based procedure described above. All predictions should be qualified with a variability range corresponding to specified confidence interval (e.g. 90% confidence). For example, the pillar is expected to fail when its width is 38 ± 23 m, or the broken ground depth is expected to be 2.2 ± 0.7 m (Fig. 21).

If the reliability of the predictions is judged unsatisfactory, then various alternatives to reducing the magnitude of C_p can be considered. It is suggested that the simplest procedures be considered first, followed by more complex alternatives in the following order:

(1) *Refine input parameters*. Firstly, attempt to minimize errors due to model input parameters. There may be a need to refine the geometric representation of our mining geometry (pillar widths, stope shapes, etc.). Although stress measurements can be used to estimate the orientation and ratio, the pre-mining stress state can be refined by back-analysis using the error table minimization procedure described in Section 3.3. A better fit may result if time under load (i.e. a time dependant strength envelope) is considered. The cost of these refinements is minimal.

(2) *Map geological complexity*. Instead of trying to model geological complexity, a simpler alternative is to determine a heterogenous rock mass strength distribution by mapping the changing geological conditions. Non-homogeneous geological details could be incorporated into the model by broadly defining different lithological units across the mine site as described in Section 5.6.1. A different strength envelope would be determined for each unit requiring calibration on a unit-by-unit basis. The cost of these refinements is repeated back-analyses for characterization of each lithological unit.

(3) *Model geological complexity*. Incorporating geological complexity by introducing zones with different stiffness or pre-mining stress states into our numerical model could be next considered. An example is where stiff dykes are well known to attract high stresses. Anisotropic behaviour of the rock mass may also be important. Note that a model with increased complexity takes more time to build, calibrate and run. Also, these refinements to the model necessitate repeated calibration back-analyses for each additional zone, and with a more complex model.

(4) *Simulate material complexity*. Finally, inelastic modelling needs consideration. The most significant contribution from inelastic modelling is stress transfer away from yielding zones. This effect can be significant in heavily mined, highly loaded areas. The rudiments of stress transfer could be incorporated into an elastic model simply by excavating failed pillars. If the rock mass response is governed by large-scale structures, fault slip simulation can be used to model important features. Here, all inelastic response is confined to the discrete fault slip surfaces. In other cases we may want to incorporate bulk rock mass yielding where a block model or a non-linear plasticity model could be used. Time-dependant strength parameters (i.e. creep response) may also have to be taken into account. The cost of these refinements is greatly multiplied owing to the increase in time required to build, calibrate and run these models. Incorporation of discrete fault slip planes requires considerable effort to define a reliable location and orientation for each structure. The error table minimization procedure (described in Section 3.3), defines a procedure for determination of appropriate rock mass constitutive parameters for inelastic models. Typically it can be anticipated that properly calibrated inelastic modelling will require one or more orders of magnitude additional effort than elastic modelling.

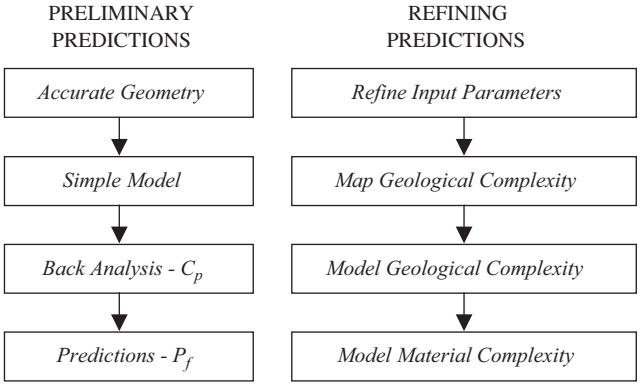

Fig. 21. Procedure for making predictions.

Each of the above refinements should be tested by comparing back-analysis results to observed in situ response. Reduction in the scatter of the clustering about the best fit would indicate that the modelling process is

performing better. This is the only definitive way of determining whether the increased effort was justified.

Unless the effort of calibration (i.e. comparing back-analysis results to observed in situ response) is made, it is unclear whether the increase in accuracy anticipated by use of a more complex model is offset by the uncertainty introduced by the additional input parameters. Without this calibration it is conceivable that one could obtain less reliable predictions or at the very least, results with unknown reliability. At the very minimum, back-analysis would be required for confirmation of model predictions before any costly decisions were made based on these. In the final analysis, all efforts at modelling and calibration can only be justified if we are able to make better design decisions. Often more questions can be answered and more information can be obtained by running many simpler models rather than a few complex ones.

In the above recommended procedure there has been little emphasis placed on detailed in situ stress measurement, laboratory testing and degradation to rock mass strength. Whilst these procedures are the only techniques for gaining information under certain circumstances, they should only be relied on when back-analysis results are not available. These procedures are most appropriate in quantifying site variability, characterizing green-field sites or anticipating changing conditions ahead of the mining face. If they are to be used, appropriate values for coefficient of variation corresponding to each of these should be combined in estimating C_p before making predictions. In all cases, the cost of using these procedures should be weighed against the benefit of conducting additional back-analyses. Many back-analyses can be completed for the same cost as an in situ stress measurement.

6. Conclusions

Operators of underground mines are routinely faced with decisions regarding both the safety of personnel and risk of ore loss under limited stability ground conditions. Ultimately a 'go–no–go' decision must be made and this is usually done using a range of design tools, monitoring and engineering judgement. One common input in this process is from numerical models that often are used to identify possible mechanisms of failure, and make relative comparisons for various design alternatives. When modelling is used to make conclusive predictions, it is essential that the modelling system performance first be tested by comparison with back-analyses under similar conditions. In this way, the uncertainty of the predictions can be estimated and clearly stated. This allows appropriate assessment of the significance of their contribution to an overall decision to be made relative to the many other factors thereby contributing to a cost effective, safe mining environment.

Reliability can be established by using statistical techniques to compare the difference of many individual predictions with their average behaviour. Subjective agreement for a few isolated cases will not suffice: well-clustered results under a wide range of conditions are required to establish the reliability of modelling predictions. Simpler elastic models have the advantage that they can be built, executed, calibrated and interpreted with minimal effort. Although in many cases inelastic models have the potential to provide more accurate predictions, they can require orders of magnitude more effort for their implementation. With realistic time constraints and a limited budget, there is clearly more value in predictions from a well-calibrated simple model, than from an inadequately calibrated complex model with unknown or poorly characterized reliability. Reliable model predictions are mandatory if designs based on these are also to be reliable. Better design decisions may be made by running many simple models rather than a few complex ones.

The observational approach is well defined and can be easily used to quantify prediction variability with a minimum of engineering effort. This need not be restricted only to stress driven failure events such as pillar failures, since even non-events (i.e. cases where no failure is observed) can be back-analysed to define a lower limit to the strength. Envelopes corresponding to a variety of stress-induced responses can be considered. This provides site-specific numbers that can be used to support engineering judgement. Where this method can be applied, it represents the best way of estimating accuracy limits since it quantifies and demonstrates the predictive capability of the entire modelling system. This includes the rock mass variability, assumptions regarding input parameters, and applicability of the chosen modelling technique.

A generalized error table minimization methodology has been described that allows field observations to be used to find the value of any desired input parameter that results in the best fit with modelling results. This can include both strength parameters and pre-mining stress state assumptions, and is applicable to both elastic and inelastic modelling. Using this procedure, the goodness of fit can be determined and the prediction reliability can be quantified.

An obvious limitation of the observational approach is that observations of rock mass response are required for its application. These are often not available in feasibility studies, "green-field" sites or in new mines. In such cases the conventional empirical method for rock mass strength estimation represents the best alternative. Efforts have been made here to quantify the uncertainty associated with this method so that the uncertainty of these predictions can be clearly stated. These estimates show that low-prediction reliability should be expected thus requiring large safety factors.

Whether there is agreement or not as to the actual magnitude of variability associated with the conventional empirical approach, one thing is clear: there is large uncertainty in the prediction accuracy obtained with this approach. More work needs to be done to firmly establish the quantitative reliability associated with this method. The large uncertainty associated with this procedure most likely

results from errors in assumptions regarding the ratio of pre-mining stress to rock mass strength. This can offset calculations from reality. This stress versus strength mismatch is not due to any limitation of the modelling capability, but rather, stems from an inability to accurately quantify the required input parameters.

High-accuracy conclusive predictions may not be possible in a geological environment. It seems unlikely that the necessary low values of C_p could ever be attained owing to the natural variability of the rock mass. If geological variability is primarily responsible for the mismatch between back-analysis results and observed in situ response, then more complex inelastic modelling will not improve prediction accuracy. Engineering effort would probably be better directed in attempting to spatially correlate the rock mass failure criterion to match the heterogenous rock mass strength distribution.

The definition of safety factor does not provide a unique measure of safety and hence should not be used to compare one situation to another on an equal basis. Even though the magnitude of safety factor can be adjusted to provide the same probability of failure depending on local site conditions, the result will still be dependent on the local value of σ_3. This introduces unnecessary uncertainty into modelling predictions that is readily avoided by use of probability methods.

The methodology proposed in this paper provides a quantitative, practical way to design underground excavations in a mining environment, with improved reliability. As a result, cost and safety related decisions could be made with a known level of confidence providing us with real numbers that can be used to backup our engineering judgement. The author believes that this work goes a long way towards improving on existing empirical procedures for design of underground excavations.

Acknowledgements

Inco is acknowledged for permitting publication of details and analysis results from Creighton Mine. The author is grateful for the comments from the formal reviewers and assistance of many colleagues who offered suggestions and constructive criticism.

References

[1] Starfield AM, Cundall PA. Towards a methodology for rock mechanics modelling. Int J Rock Mech Min Sci 1988;25:99–106.
[2] Hoek E. The challenge of input data for rock engineering. ISRM New J 1994;2:23–4.
[3] Harr ME. Reliability based design in civil engineering. New York: McGraw-Hill Book Company; 1987.
[4] Cai M, Kaiser PK, Uno H, Tasaka Y, Minami M. Estimation of rock mass deformation modulus and strength of jointed hard rock masses using the GSI system. Int J Rock Mech Min Sci 2004;41:3–19.
[5] Martin CD. Seven years of stress measurements at the URL—an overview, Rock mechanics contributions and challenges. In: Proceedings of the 31st US Symposium. Colorado: Golden; 1990. p. 15–26.
[6] Villaescusa E, Li J. A review of empirical methods used to estimate rock mass compressive strength and deformability in the mining industry. MassMin Chile; 2004.
[7] Hoek E, Brown ET. Practical estimates of rock mass strength. Int J Rock Mech Min Sci 1997;34:1165–86.
[8] Ramamurthy T. Stability of rock mass. Indian Geomech J 1986;16:1–74.
[9] Trueman R. An evaluation of strata support technique in dual life gate roads. PhD. Thesis, University of Wales, Cardiff, 1988.
[10] Singh B. Indian case studies of squeezing grounds and experiences of application of Barton's Q-system. In: Workshop on Norwegian method of tunnelling. New Delhi: CSMRS; 1993.
[11] Kalamaras GS, Bieniawski AT. A rock mass strength concept for coal seams incorporating the effect of time. In: Proceedings of the eighth international congress on rock mechanics, vol. 1. Rotterdam, A.T.: ISRM, AA Balkema; 1995. p. 295–302.
[12] Sheorey PR. Empirical rock failure criteria. Rotterdam: AA Balkema; 1997.
[13] Hoek E. Practical rock engineering. In: Hoek E, editor. Course notes. Canada: Vancouver; 1998.
[14] Martin CD, Maybee WG. The strength of hard-rock pillars. Int J Rock Mech Min Sci 2000;37:1239–46.
[15] Hoek E. Reliability of Hoek–Brown estimates of rock mass properties and their impact on design. Int J Rock Mech Min Sci 1998;35:63–8.
[16] Kayabasi A, Gokceoglu C, Ercanoglu M. Estimating the deformation modulus of rock masses: A comparative study. Int J Rock Mech Min Sci 2003;40:55–64.
[17] Bieniawski ZT. Determining rock mass deformability: Experience from case histories. Int J Rock Mech Min Sci 1978;15:237–47.
[18] Asef MR, Reddish DJ, Lloyd PW, Mitri. Rock support interaction analysis based on numerical modeling. Geotech Geol Eng 2000;18(1):23–37.
[19] Serafim JL, Pereira JP. Considerations on the geomechanical classification of Bieniawski. In: Proceedings of the symposium on engineering geology and underground openings. Portugal: Lisboa; 1983. p. 1133–44.
[20] Nicholson GA, Bieniawski ZT. A non-linear deformation modulus based on rock mass classification. Int J Min Geol Eng 1990;8: 181–202.
[21] Hudson JA, Cooling CM. In situ rock stresses and their measurement in the UK—Part I. The current state of knowledge. Int J Rock Mech Min Sci 1988;25:363–9.
[22] Wiles TD, Kaiser PK. A new approach for statistical treatment of stress tensors. Specialty Conference on Stress. Ottawa, Canada, 1990.
[23] Martin CD, Kaiser PK, Christiansson R. Stress, instability and design of underground excavations. Int J Rock Mech Min Sci 2003;40:1027–47.
[24] Wiles TD. Accuracy and applicability of elastic stress analysis methods in mining. In: 15th Canadian rock mechanics symposium, Toronto, Canada, 1988, p. 207–220.
[25] Wiles TD. Map3D user's manual, Mine modelling (Pty) report, 2005.
[26] Lipson C, Sheth NJ. Statistical design and analysis of engineering experiments. New York: McGraw-Hill Book Company; 1973.
[27] Terzaghi K, Peck RB. Soil mechanics in engineering practice. New York: Wiley; 1967.
[28] Jing L. A review of techniques, advances and outstanding issues in numerical modelling for rock mechanics and rock engineering. Int J Rock Mech Min Sci 2003;40:283–354.
[29] Marisett SD. Assessing pillar design at Inco's Creighton mine using the local energy release density method, MASc. Thesis, Queen's University, Kingston, Canada, 2001.
[30] Wiles TD, Marisett SD, Martin CD. Correlation between local energy release density and observed bursting conditions at Creighton mine. mine modelling report, Sudbury, Canada, 1998.
[31] Wiles TD. Rockburst prediction using numerical modelling: Realistic limits for failure prediction Accuracy. In: Sixth international symposium on rockbursts and seismicity in mines (RaSiM 6), Perth, Australia, 2005.

[32] Wiles TD, Kaiser PK. In situ stress determination using the under-excavation technique—I. Theory. Int J Rock Mech Min Sci 1994; 31:439–46.
[33] Wiles TD, Kaiser PK. In situ stress determination using the under-excavation technique–II. Applications. Int J Rock Mech Min Sci 1994;31:447–56.
[34] Wiles TD, Villaescusa E, Windsor CR. Rock reinforcement design for overstressed rock using three-dimensional numerical modeling. In: Fifth international symposium on ground support in mining and underground construction (ground support 2004). Australia: Perth; 2004.
[35] Daehnke A, Andersen LM, De Beer D, Esterhuizen GS, Glisson FJ, Grodner MW, et al. Stope face support systems, SIMRAC GAP330 project report, Johannesburg, 1998.
[36] Rosenbleuth E. Two-point estimates in probabilities. J Appl Math Modelling 1981;5:329–35.
[37] Obert L, Duval WI. Rock mechanics and the design of structures in rock. London: Wiley; 1967.
[38] Hoek E, Bray JW. Rock slope engineering, Installation. London: Mining and Metallurgy; 1977.
[39] Thomas HH. The engineering of large dams. London: Wiley; 1976.
[40] Nicholls D. personal communication, 1992.
[41] Lachenicht R, Wiles TD, van Aswegen G. Integration of deterministic modelling with seismic monitoring for the assessment of rock mass response to mining: Part II applications. In: Fifth international symposium on rockbursts and seismicity in mines (RaSiM 5). South Africa: Sandton; 2001. p. 389–96.
[42] Wiles TD, Lachenicht R, van Aswegen G. Integration of deterministic modelling with seismic monitoring for the assessment of rock mass response to mining: Part I theory. In: Fifth International symposium on rockbursts and seismicity in Mines (RaSiM 5). South Africa: Sandton; 2001. p. 379–88.