Slope Stability 2013 – P.M. Dight (ed)

doi:10.36487/ACG_rep/1308_19_Thomas

A statistical approach to account for elevated levels of uncertainty during geotechnical design

R.D.H. Thomas *Coffey Mining Pty Ltd, Australia*

Abstract

Uncertainty in geotechnical engineering results from the inherent variability of natural materials, and the challenges engineers and geologists have in correctly assessing them. Geotechnical design at the definitive feasibility study (DFS) stage should be based on a geotechnical model with target levels of confidence of ideally between 50 and 75%. Due to various reasons such as budgetary constraints for ground investigation or changes to the resource model affecting mine design, a higher degree of data uncertainty can exist in one or many of the model inputs. This can be accounted for by selection of conservative design values, but where probabilistic design is desired further uncertainty can be induced by estimating population characteristics, for instance by adopting coefficients of variation from published literature, e.g. Harr (1987) and Kim (2005).

This paper presents a case study for a DFS level open pit slope design for a gold project in West Africa. The geological, structural and hydrogeological models were suitably defined, however limited geotechnical drillhole data was available for some units. The limited data sets hindered rock mass characterisation and derivation of design values (and distributions) for subsequent slope stability modelling, resulting in elevated levels of uncertainty. The author was faced with the prospect of either accepting the uncertainty and accounting for the small populations by choosing lower bound values and assumed distributions, or relying on regional data from experience of working with the encountered units elswhere. In an effort to ensure the most relevant, deposit specific data was used, the author sought to supplement the limited data sets with additional drillcore data from elsewhere within the project area.

The challenge the author faced was to justify combining data that would conventionally be considered separately. A number of statistical tests were used to demonstrate the validity of combining the different populations of drillcore data. A suite of logged and derived parameters were tested. The results of the statistical analyses and the effect of combining drillcore data populations on the resultant level of uncertainty and ultimate pit slope design are presented.

1 Introduction

Uncertainty in geotechnical engineering results from the inherent variability of natural materials, and the challenges engineers and geologists have in correctly assessing them.

During all stages of a geotechnical design a degree of uncertainty will exist. Acceptable levels of uncertainty or data confidence at different project stages have been defined and related to levels of geotechnical effort in guidelines such as CANMET's pit slope manual (CANMET, 1976) and those produced by CSIRO (Read and Stacey, 2009). The CSIRO guidelines state that geotechnical design at the definitive feasibility study (DFS) stage should be based on a geotechnical model with target levels of confidence of between 50 and 75%. This is made up of target levels of confidence of the inputs to the geotechnical model as follows:

- Geological 65–85%
- Structural 45–70%
- Hydrogeological 40–65%
- Rock mass 60–75%

Due to various reasons such as budgetary constraints for ground investigation or changes to the resource model affecting mine design, a higher degree of data uncertainty can exist in one or many of the model inputs. The geological, structural and hydrogeological models are often prepared and presented to the geotechnical engineer, with communication of the associated levels of confidence. These models are likely to be validated against results of specific geotechnical investigations where possible, and uncertainty or alternative interpretations incorporated into the design process. Little can be done to reduce uncertainty – it must be accounted for, however, by undertaking sensitivity analyses, adjusting acceptance criteria (Harr, 1987), or using reliability based optimisation (Lilly, 2000).

The rock mass model is most often constructed from drillhole derived logging and laboratory test data. While the spacing of drillholes undertaken to define the resource and construct the geological model is strictly governed by industry reporting codes, the required spacing of geotechnical drillholes is much less rigorously defined. While the required drillhole spacing will vary from deposit to deposit, the drillhole programme should be designed with drillholes spaced to allow the various rock mass units within the geotechnical model to be suitably characterised for the level of study being undertaken.

From DFS level design upwards, it is generally desired that geotechnical design is undertaken using probabilistic, as well as deterministic methods. In order to achieve this both a characteristic value and a distribution of design parameters from the rock mass model are required. If the distributions are not well defined, further uncertainty can be induced by estimating population characteristics, for instance by adopting coefficients of variation from published literature, e.g. Harr (1987) and Kim (2005). The rock mass model tends to be the input to the geotechnical model for which the geotechnical engineer has the most responsibility and greatest degree of influence, and is the focus of the case study presented below.

2 Case study

2.1 Background

Geotechnical design for two deposit areas (A and B) was to be undertaken at a gold project in West Africa which progressed directly from a Scoping level study to DFS, without a pre-feasibility level of study. The two deposits were structurally hosted and directly along strike of each other, approximately 1,200 m apart. The geotechnical investigation involved the drilling of a number of drillholes in each deposit, however due to the early onset of the wet season, only one geotechnical drillhole could be collared in the footwall of deposit B. The geological, structural and hydrogeological models were provided by others. A simplified geological map of the deposits, along with the collar locations of the geotechnical drillholes is shown in Figure 1.

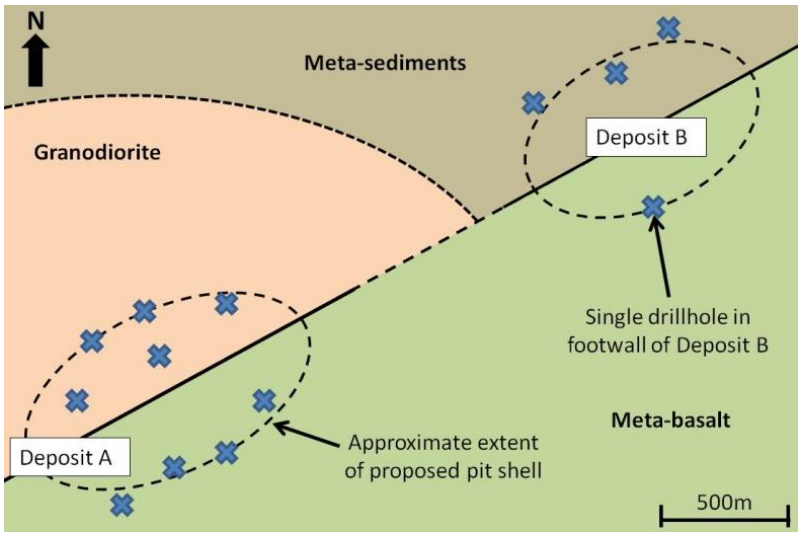

Figure 1 Simplified geological map showing the locations of deposits A and B, the mineralised structure, and the drillhole collar locations

2.2 Definition of the problem

The footwall of both deposits consisted of the same geological unit (meta-basalt) whereas the hangingwall varied, consisting of granodiorite in deposit A, and meta-sediments in deposit B. During the geotechnical design process, subdivision of drillhole data by geotechnical domain and weathering grade led to limited sample populations for the weathered and fresh material in the footwall of deposit B, where data was available from just the single drillhole. The weathered domain was assessed to be of limited thickness (<20 m) and not considered to significantly affect serviceability. This limited data set of the fresh material however, was considered likely to hinder rock mass characterisation and derivation of design values and distributions for subsequent slope stability modelling. The resultant level of uncertainty was considered unlikely to meet that required at a DFS level, and the author faced the prospect of either:

- Using the limited data set, accepting the residual uncertainty, and accounting for the small populations by choosing lower bound values (and assumed distributions).
- Relying on regional data from previous experience working with the encountered units.
- Supplementing the limited data set with additional drillcore data from within the footwall of deposit A.

Neither of the first two options was particularly palatable to the author due to the sensitivity of the project to strip ratio (i.e. slope angle) and the desire to limit uncertainty and undertake probabilistic design at DFS level. The third option would require the combining of data from a distance of 1,200 m from the deposit, all be it along strike and in a similar structural and geological setting.

The option of supplementing the information for deposit B with additional drillcore data from deposit A was preferred, however, without testing the validity of combining the data sets, this approach was considered difficult to apply with confidence.

In order to test the validity of combining the data sets, a method of comparing the two data sets is required. Data sets of drillcore data consists of both logged interval and point data, and both categories include qualitative and quantitative parameters. Generally, the qualitative parameters, such as rock type, and weathering grade are used for subdivision of drillhole data and the design process seeks to characterise each of these groups.

The quantitative logged parameters are generally assessed statistically to allow the selection of representative values for rock mass characterisation and so statistical testing was considered. As different logged parameters can exhibit markedly different distributions, it was understood that a number of different statistical tests may be required.

The following sections outline the approach used for statistically testing the validity of combining the data sets derived from drillcore.

2.3 Statistical testing

To combine data sets, a method for testing whether the two populations of data are from the same parent populations (in this case a single rock mass unit) is required. In statistical terms, we are required to test a hypothesis, that being that the two data sets are from the same parent populations i.e. are equivalent. This is the ‘null hypothesis’ (that the difference between the parent population is zero), and can be justified by the apparently consistent geological and structural setting of the footwall of both deposits. The ‘alternative hypothesis’ is that the two data sets have been sampled from parent populations that are not equivalent. The statistical tests considered compare the calculated difference between the two data sets with the difference that would be expected were the samples within the two data sets randomly selected from the parent population. A confidence interval is selected at which the test is performed, and this confidence interval represents a probability at which the ‘null hypothesis’ is rejected. For instance a confidence interval of 95%, translates to a 95% probability that the true value lies within this range. A value outside of this

range is expected to occur at a probability of 5%, if sampling is random. Should comparison of the two data sets result in a value outside of this range, then grounds to reject the 'null hypothesis' are met.

A population of data can generally be defined by a measure of central tendency (often the mean or median) and a measure of spread (the standard deviation, or variance). In order to test data sets and have justification for combining them, the equivalence of both of these characteristics needs to be demonstrated. Normally distributed data can be tested using parametric tests, which include a widely known suite of statistical methods, including the Student t-test, for the testing of the equivalence of the mean, and the F-test, for testing the equivalence of variance. For these tests to be valid, it must be assumed that the parent populations have distributions whose shapes are normal, or approximately normal. Due to a number of logged drillcore parameters often exhibiting non-normal distributions, as discussed above, alternative statistical tests, known as non-parametric tests, also need to be considered and are discussed below. These tests have the advantage that they do not require normality (or near normality).

2.3.1 Parametric tests

2.3.1.1 Testing equivalence of the mean - Student t-test

The Student t-test was developed by William Gosset, while working at the Guinness brewery in the early 1900s. The test compares the difference between the means of two populations in relation to the variation in the data (expressed as the standard deviation of the difference between the means). The t-test will calculate the t-statistic, *t*, using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2 + s_2^2)/n}} \quad (1)$$

Where:

$\bar{x}_1$ = the mean of the data within group 1.

s_1 = the standard deviation of group 1.

n = the number of samples (total of both groups).

A 'null hypothesis', that the means of the two populations do not differ, is tested, at a defined confidence level (generally 95%). The resultant *t* value can be equated to a probability (P value) based on the degrees of freedom (related to the total number of samples) and a conclusion drawn as to whether the sample means are equivalent or not.

The Student t-test uses the Student t-distribution, which is based on the normal distribution, and is strictly only valid when the populations under consideration are normally distributed. Using the Central Limit Theorem, however, the Student t-test can be used provided 'our sample size is large and the population does not differ too much from normality' (Davis, 2002).

2.3.1.2 Testing equivalence of the variance – F-test

To test whether the variance of two samples is statistically similar the F-distribution can be used. The F−distribution is the theoretical distribution of values that would be expected by randomly sampling from a normal population and calculating the ratio of variances for all possible pairs of samples, where the F-ratio is as follows:

$$F = s_1^2/s_2^2 \quad (2)$$

Where:

s_1 = the standard deviation of the data within group 1 (the variance being the square of this).

The F-test works in a similar way to the Student t-test, in that the 'null hypothesis,' is achieved when the sample populations have equivalent variances, at a chosen significance level. This test also requires that the parent population is normal. Figure 2 (after Davis, 2002) shows a typical F-distribution. In the example shown the 95% confidence interval relates to an expected value of F of 2.24, the value above which would result in rejection of the 'null hypothesis.'

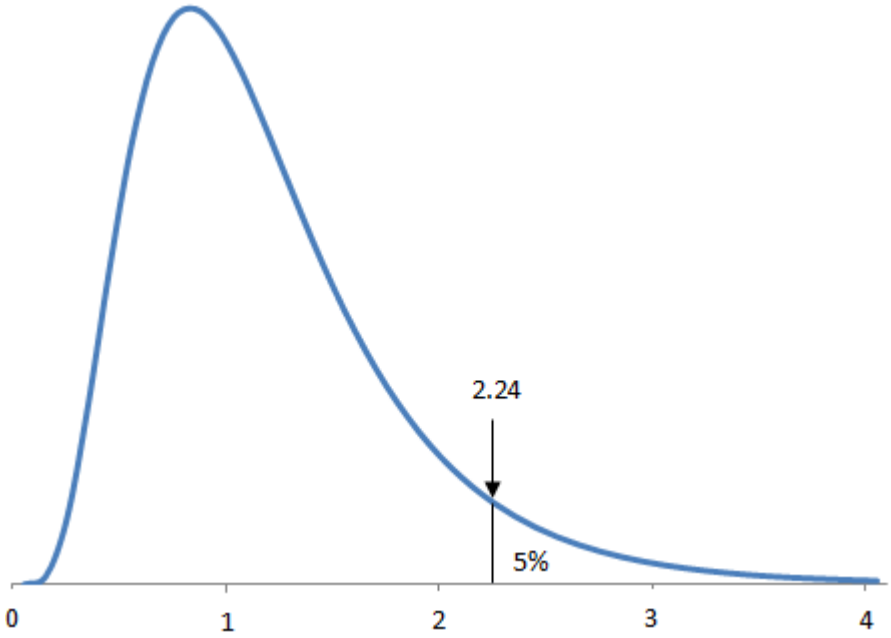

Figure 2 An example F-distribution, showing a critical value of F = 2.24, corresponding to a confidence interval of 95% (after Davis, 2002)

2.3.2 Non-parametric tests

For some logged parameters, including Rock Quality Designation (RQD) and fracture spacing, the distributions tend to be highly non-normal. In these cases the Student t and F-tests cannot be applied, and non-parametric statistical testing is required. The non-parametric equivalent to the Student t-test and the F-test are the Mann–Whitney test and the Brown–Forsythe test. Both require that the two samples under consideration have similarly shaped distributions (negating the need to test the equivalence of skew), but can be non-normal. These non-parametric tests can also be used on normal distributions.

Non-parametric tests, as discussed in Davis (2002) use information of a lower rank than parametric tests, such as nominal or ordinal observations. The median (or median rank) rather than the mean value is assessed as the measure of central tendency. As a result, non-parametric tests have the added advantage that they are less likely to be adversely affected by outliers.

2.3.2.1 Testing equivalence of the median - the Mann–Whitney test

The Mann–Whitney test is the non-parametric equivalent of the Student t-test. The test compares the medians of the two samples in a similar way that the Student t-test compares means. Ranking of the combined populations of the two data sets under consideration is undertaken, and the sum of the ranks of each populations compared against a confidence interval, with a critical value taken from a distribution (equivalent to the Chi distribution, where n >8). A detailed explanation of the Mann–Whitney test (also closely related to the Kruskal-Wallis test) is given in Davis (2002).

2.3.2.2 Testing equivalence of the variance – the Brown–Forsythe test

The Brown and Forsythe (1974) test is the non-parametric equivalent of the Mann–Whitney test. The test compares the absolute variation from the medians of the two samples, and is based on Leven's test which

compares the deviation to the mean. The Brown–Forsythe version of the test was developed by Brown and Forsythe (1974) and is generally preferred when the distribution of the underlying data is unknown.

2.3.3 Applicability to drillcore data

The aforementioned statistical tests all rely on the premise that the data points within each sample are selected at random from the parent population. When data points are selected from a single drillhole within a domain it is likely, due to the spatial variation of parameters, that the intervals sampled and logged are not a random selection of the parent population. This suggests that strict application of the tests may be inappropriate.

In order to assess the effects of sampling from a single drillhole on the above mentioned tests, data for a single rock mass unit derived from a number of drillholes within a single geotechnical domain can be tested and the results of tests compared. The drillcore data from the fresh zone of the footwall of Deposit A was selected for this testing, a data set consisting of data from four drillholes. The test process involves removing a single drillhole from the data set and comparing the resulting sub-populations, one a single drillhole, and one the remaining three drillholes. As the original data set is known to comprise of data from a single geotechnical unit, i.e. the same parent population, the results of this preliminary testing should demonstrate the effect of spatial variability on the test results, for the various parameters considered.

2.4 Selection of drillcore derived parameters and appropriate tests

During geotechnical drillcore logging a number of parameters are generally recorded, including the weathering grade, the field index strength, the RQD, the fracture spacing, and the orientation and condition of discontinuities. Of these a number are quantitative or not likely to be reflective of the rock mass condition. As a result the following logged parameters have been selected for analysis:

- Estimated field index strength.
- RQD.
- Fracture spacing.

In addition to these logged parameters the following derived parameters, routinely calculated in the design process, are also to be assessed:

- Rock Mass Rating$_{89}$ (RMR$_{89}$; Bieniawski, 1989) – a rating system used to characterise the rock mass from drill core data.
- Joint Roughness Coefficient (JRC; Barton and Choubey, 1977) – based on recorded roughness and planarity data.

The first step of undertaking the analysis involves the plotting of histograms of all the test parameters for each of the parent data populations. This was undertaken to allow selection of the most suitable testing methodology. The distributions of the whole data set for the footwall of deposit A, fresh material is shown in Figure 3.

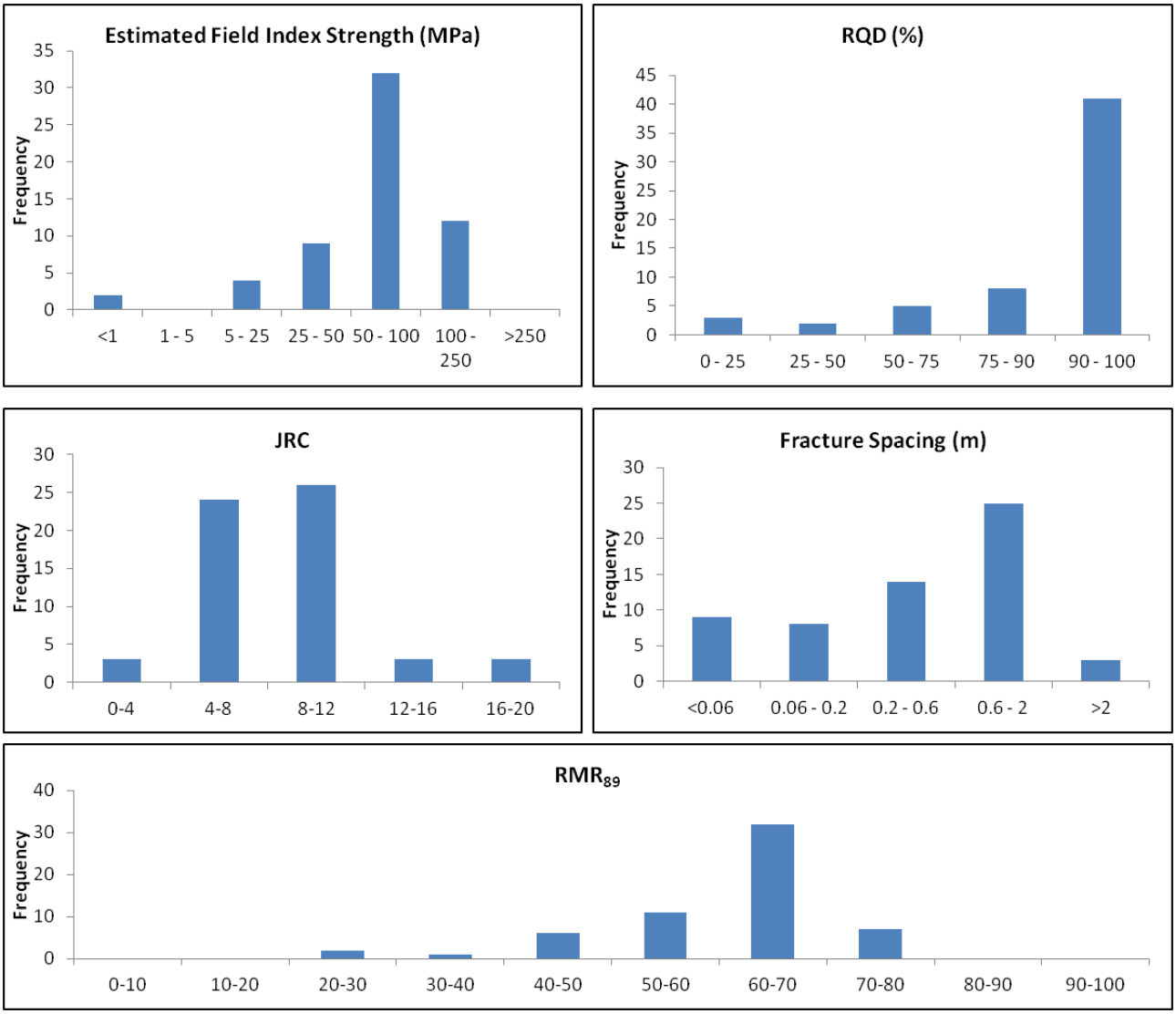

Figure 3 Histograms of the drillcore data from the fresh zone of the footwall of deposit A. The data set consists of data from four drillholes

Of the parameters being considered for testing, the distributions of RQD, fracture spacing and estimated field index strength (note the varying bin sizes) are evidently non-normal, and require non-parametric testing. Of the JRC and RMR_{89}, the JRC is noted to contain values at the extremes of the available range and hence would also be better assessed by non parametric tests. RMR_{89} on the other hand can be assessed parametrically, as the distribution is close to normal. Non-parametric tests could also be undertaken on the RMR_{89} should this be desired.

Each test will be carried out on five drillcore derived parameters, and both the measure of central tendency and the spread will be assessed.

2.5 Results of preliminary testing and discussion

It was decided to undertake testing of all the drillholes at both the 90% and 95% confidence intervals to allow assessment of which interval was more appropriate. The results of the testing are shown in Table 1 and Table 2 at the 90% and 95% confidence intervals respectively. The results are presented for the four tests conducted, one with data from each of the four drillholes compared against the remaining three. Ticks indicate that the statistical tests were passed, and crosses indicate failed statistical tests (i.e. the 'null hypothesis' was rejected).

Table 1 Results of testing at the 90% confidence interval

Test Number	1		2		3		4	
Measure	Central Tendency	Spread	Central Tendency	Spread	Central Tendency	Spread	Central Tendency	Spread
Field index strength	✗	✓	✓	✓	✓	✓	✓	✓
RQD	✓	✓	✓	✓	✓	✓	✓	✓
Fracture spacing	✓	✓	✓	✓	✓	✓	✓	✓
JRC	✓	✓	✗	✓	✓	✓	✓	✓
RMR_{89}	✓	✓	✓	✗	✓	✗	✓	✓

Table 2 Results of testing at the 95% confidence interval

Test Number	1		2		3		4	
Measure	Central Tendency	Spread	Central Tendency	Spread	Central Tendency	Spread	Central Tendency	Spread
Field index strength	✗	✓	✓	✓	✓	✓	✓	✓
RQD	✓	✓	✓	✓	✓	✓	✓	✓
Fracture spacing	✓	✓	✓	✓	✓	✓	✓	✓
JRC	✓	✓	✓	✓	✓	✓	✓	✓
RMR_{89}	✓	✓	✓	✗	✓	✗	✓	✓

It can be seen from the results presented in Tables 1 and 2, that despite the data all being from a single rock mass unit, a number of failures were noted. These failures occurred on tests undertaken on all but RQD and fracture spacing. A maximum of two tests were failed when comparing drillhole 2 against the remaining drillholes at 90% confidence interval. Generally the results at both the 90% and 95% confidence intervals were similar, with one additional failure in tests conducted at the 90% confidence interval.

At both confidence intervals only test 4 resulted in no rejections of the 'null hypothesis', despite all the data coming from parts of a rock mass unit conventionally considered as one. The noted failures of statistical tests in test numbers 1, 2, and 3, are attributed to the non-random nature of the sampling when all data comes from the one drillhole. It appears that the variability present within a rock mass unit can result in individual parameters failing statistical testing if limited data sets of only one drillhole are used.

At a confidence interval of 95%, three of the four tests had one parameter fail for either central tendency or spread. None of the four tests resulted in failures of more than one parameter, and it appears to be unlikely for multiple parameters to fail statistical testing for different data sets taken from within a single rock mass unit.

2.6 Methodology for testing the validity of combining drillcore data

From the results of the preliminary testing presented above, the following method was developed to test the validity of combining the drillcore data from the footwall of the deposits B with the corresponding data from deposit A:

- Undertake the testing on all of the logged and derived drillcore parameters, for both the measure of central tendency and the spread.
- When undertaking the testing at a confidence interval of 95%, the results of all tests should be considered as follows:
    - Failure of more than one test indicates that the combining of the data cannot be undertaken with confidence.
    - Failure of one test indicates that the combining of data is likely to be justified, but should be undertaken with caution.
    - Failure of no tests indicates that the data can be combined with a high level of confidence.

2.7 Testing the validity of combining the data from deposits A and B

The distributions of the logged and derived drillhole parameters from the fresh material in the footwall of deposit B were compared to the distributions in Figure 2, and the distributions were assessed to be consistent. The statistical tests as discussed above were applied to the data at the 95% confidence interval, and the results are presented in Table 3.

Table 3 Results of testing at the 95% confidence interval for data from deposits A and B

Measure	Central Tendency	Spread
Field index strength	✓	✓
RQD	✓	✓
Fracture spacing	✓	✓
JRC	✓	✓
RMR_{89}	✓	✓

The results of the statistical testing were that all of the tests were passed. The 'null hypothesis,' that the parent populations of the data were equivalent, could not be rejected for any of the individual tests. As this only occurred in one of the four tests undertaken on data from drillholes from with deposit A, the test results demonstrate that the data sets can be combined with a high degree of confidence.

2.8 Effect on rock mass characterisation and slope design

The combining of the data sets resulted in an increase in the size of the data set for the fresh material in the footwall of deposit B. The data set increased from 19 data points to 78. The resulting increase in confidence / reduction in uncertainty affected the choice of representative rock mass parameters for use in design. Table 4 demonstrates the effect of combining the data sets on the representative values taken forward to design by considering the original data set alongside the augmented data set. The design values were chosen with regard to the amount of data within, and shape of the data set available.

Table 4 Rock mass characterisation for design

	Original Data Set (N = 19)		Combined Data Set (N = 78)	
Measure	**Design Value**	**Reasoning**	**Design Value**	**Reasoning**
Field index strength	15 MPa	Mean minus one standard deviation	35 MPa	Mean value
RMR_{89}	53	Mean minus one standard deviation	61	Mean value
Geological strength index	48	$RMR_{89} – 5$	56	$RMR_{89} – 5$
Coefficient of variation	30 %	From literature, insufficient data to define	17%	Derived from RMR_{89} distribution

The above data shows a marked contrast in the design values between the original and augmented data sets. In all of the parameters carried forward for design, the expanded data set allowed the justifiable selection of parameters that related to a better defined, and in this case, more competent, rock mass.

The resultant effect of the selection of these material parameters on the slope design has been assessed using limit equilibrium analysis, assuming rock mass failure only. Based on a pit depth of 130 m and equivalent acceptance criteria, an increase of overall slope angle of approximately 5° was achieved using the combined data set. Although not the only consideration in formulating the ultimate slope design, this highlights the potential benefits of this method.

3 Conclusions and limitations

The case study presented above has successfully shown a method for utilising statistical tests when assessing the similarities of drillcore data from different deposit areas. The method tests the hypothesis that the data is of the same parent population, and to consider undertaking the tests a geological argument for combining the data sets should first exist. Both parametric and non-parametric tests have been used and should be selected based on the data distribution of the logged parameter being considered.

Statistical testing relies on data randomly selected from a parent population, a case that may not be met when data is all from a single drillhole due to spatial variation of the data within a unit. The preliminary testing of subsets of data derived from multiple boreholes in the one unit allows the spatial variation of parameters to be accounted for. The results of the preliminary testing were assessed and allowed a methodology to be developed for testing the validity of combining drillcore data from different deposit areas. The test methodology for this case study was as follows:

- Undertake the statistical testing of the two data sets on all of the logged and derived drillcore parameters (estimated field index strength, RQD, fracture spacing, JRC and RMR_{89}), for both the measure of central tendency and the spread.
- Testing at a confidence interval of 95 %, the results of all tests should be considered thus:
  - Failure of more than one test indicates that the combining of the data cannot be undertaken with confidence.
  - Failure of one test indicates that the combining of data is likely to be justified, but should be undertaken with caution.
  - Failure of no tests indicates that the data can be combined with a high level of confidence.

The degree of spatial variation is considered to be likely to vary from unit to unit, and the preliminary testing of data from a single unit is seen as an integral part of the test methodology.

While this method allowed derivation of representative material parameters and distributions, it does not replace the need for collection of geotechnical data. This method does not replace deposit specific information on the minor and major structures within the rock mass. It is suggested that where this method is used to increase confidence during slope design, the confirmation of rock mass conditions during implementation becomes even more critical.

Acknowledgement

The author acknowledges Dr Sanjive Narendranathan for his comment and advice, and Coffey International Ltd for providing the resources to publish this document.

References

Barton, N.R. and Choubey, V. (1977) The shear strength of rock joints in theory and practice, Rock Mechanics, Vol. 10(1–2), pp. 1−54.

Bieniawski, Z.T. (1989) Engineering rock mass classifications, John Wiley & Sons, Inc., New York, 251 p.

Brown, M.B. and Forsythe, A.B. (1974) Robust tests for equality of variances, Journal of the American Statistical Association, American Statistical Association, Vol. 69, pp. 364–367.

CANMET (1976) Pit slope manual, 10 chapters, CANMET, Ottawa, Canada.

Davis, J.C. (2002) Statistics and data analysis in geology, 3rd edition, Wiley & Sons, New York, 677 p.

Harr, M.E. (1987) Reliability based design in civil engineering, McGraw Hill, London, 303 p.

Kim, H. (2005) Spatial variability in soils: stiffness and strength, PhD Thesis, Georgia Institute of Technology, Atlanta, USA, August 2005.

Lilly, P.A. (2000) The minimum total cost approach to optimum pit slope design, in Proceedings International Symposium on Mine Planning and Equipment Selection, 6–9 November 2000, Athens, Greece, A.A. Balkema, Leiden, pp. 77–81.

Read, J.R.L. and Stacey, P.F. (2009) Guide for open pit slope design, 1st edition, CSIRO, Collingwood, Australia, 496 p.