Slope Stability 2020 - PM Dight (ed.)

doi:10.36487/ACG_repo/2025_37

Bayesian approach for the assessment of sufficiency of geotechnical data

LF Contreras *SRK Consulting, Australia*

M Serati *The University of Queensland, Australia*

DJ Williams *The University of Queensland, Australia*

Abstract

The characterisation of geotechnical materials for the design of mine and civil slopes requires the collection of data through site and laboratory investigations. The information provided by data contributes to the reduction of the knowledge uncertainty of design parameters. However, the amount of data collected at different stages of project development is normally limited and subjected to budget and time constraints. Proper assessment of data sufficiency at each stage is, therefore, a key aspect of the slope design process. The paper discusses the common technique used to relate the number of data points with safe values of the design parameters based on the concept of confidence interval (CI) from classical statistics (i.e. the frequentist approach). This conventional approach is then contrasted with a technique based on the highest density interval (HDI) from Bayesian statistics, which offers a simpler and more intuitive way to judge the sufficiency of data. The discussion is illustrated with examples of analysis of uniaxial compressive strength (UCS) data, and the intact rock strength parameters σ_{ci} and m_i of the Hoek–Brown strength criterion.

Keywords: *uncertainty, Bayesian statistics, data quantity, parameter confidence*

1 Introduction

One of the most important aspects of the slope design process should be the determination and reporting of the uncertainties in the various components of the geotechnical model to ensure that the uncertainty levels are commensurate with the stages of project development (Read 2009). This is often poorly undertaken. There is a general understanding of the causes and consequences of geotechnical uncertainty in the slope design process and several methods are available to quantify uncertainty. The aspect that still requires better treatment is reporting and relating the target levels of uncertainty at various project stages to the amount of data required to support the design. Traditionally, the relationship between data quantity and the uncertainty of design parameters has been treated within the framework of classical statistical methods, also known as frequentist methods, using the concept of confidence interval (CI) (Gill et al. 2005). In the frequentist approach, data is considered as a random variable whereas the model parameters sought with the support of data are assumed to be fixed unknown quantities. These considerations are the cause of various drawbacks of the methods based on the CI to assess the sufficiency of data. In contrast, the Bayesian approach of statistical analysis turns the problem around so that the parameters are considered uncertain variables while the data is assumed to be a fixed quantity. In this case, the highest density interval (HDI) is used to quantify the parameter uncertainty and to relate it to the amount of data in a way that offers advantages for the assessment of sufficiency of data.

This paper initially presents a brief description of the approaches to quantify uncertainty with emphasis on the contrast between the frequentist and Bayesian approaches of statistical analysis. The discussion of the topic of sufficiency of data includes the frequentist perspective based on the use of the CI as described by Gill et al. (2005) with an example of the analysis of uniaxial compressive strength (UCS) data to illustrate the drawbacks of the approach. Finally, the Bayesian perspective on the sufficiency of data is presented including the examples of analysis of UCS data, and the intact rock strength parameters σ_{ci} and m_i of the Hoek–Brown strength criterion.

2 Probabilistic approaches to quantify uncertainty

There are two main approaches of statistical analysis known as frequentist (or classical) and Bayesian. These methodologies refer in particular to statistical inference analysis where data is used to draw conclusions on the characteristics of the population represented by the data. The objects of the inference analysis are, therefore, the parameters used to describe the population. This process has uncertainty, which is measured with probability values. The conceptual basis of the two approaches differ in terms of what is considered uncertain (data or parameters), and on the interpretation of probability (VanderPlas 2014).

2.1 The classical approach for inference of parameters

The frequentist approach is based on the concept of data, which is used to characterise the population from which it is drawn, as being the result of a random sampling process. Therefore, in this approach data is considered uncertain whereas the parameters investigated are unknown fixed quantities. In this case, probabilities are interpreted as relative frequencies of outcomes from randomised trials or samples. Meaningful probabilities require to be based on numerous trials; hence, it is implicit in the approach that many samples (data) are necessary for an accurate characterisation of the population. Frequentist statistical methods are used by default in many areas of engineering design, including the geotechnical design of mine slopes; however, the implications of the conceptual basis are often unclear to the analyst, leading to misinterpretation of results.

The results of the inference analysis of parameters consist of point estimates (e.g. the mean) and error measures (e.g. the confidence interval) of the parameters investigated. The CI defines the range into which the population mean, also called the true mean, lies and it is defined as follows:

$$\overline{x} - t_{(\alpha,n-1)} \frac{CV\,\overline{x}}{\sqrt{n-1}} \leq true\ mean \leq \overline{x} + t_{(\alpha,n-1)} \frac{CV\,\overline{x}}{\sqrt{n-1}} \tag{1}$$

where:

$\bar{x}$	=	the mean of the data points.
$t_{(\alpha,n-1)}$	=	the confidence coefficient obtained from the Student-t distribution with (n - 1) degrees of freedom for a confidence level equal to $(1-2\alpha)$.
n	=	the number of data points.
CV	=	the coefficient of variation of the dataset.

Therefore, for a confidence level of 95% α = 2.5%. The CIs calculated in this way are said to correspond to the small sampling theory, which typically corresponds to situations where the number of data points is less than 30. The Student-t distribution gets very close to the normal distribution as the number of degrees of freedom increases and in general, the *t* factor in Equation 1 tends to a constant value for more than 30 data points, which corresponds to the large sample theory situation.

Geotechnical information is normally characterised by small datasets and the CI as described in Equation 1 is used for the assessment of the confidence of data. Gill et al. (2005) describe a procedure for estimating the number of data points required to define rock properties with a specified level of confidence using the CI from the small sampling theory.

2.2 The Bayesian approach for inference of parameters

In the Bayesian approach, data is combined with the existing prior knowledge on the parameters investigated into a so-called posterior distribution using the Bayes rule. In this case, data represents a particular state of information on the population and therefore are considered fixed, whereas the parameters sought to characterise the population are uncertain and represented by random variables. The posterior distribution indicates the likelihood and correlations of the parameters investigated with the

Bayesian analysis, providing a measurement of their uncertainty. These results are specific to the state of knowledge included in the data and the priors used in the evaluation. Probabilities within a Bayesian framework are interpreted as 'degrees of belief' that can be assigned directly to situations or events. The posterior distributions are normally complicated functions that require special methods of evaluation such as the Markov Chain Monte Carlo (MCMC). Detailed information on the Bayesian approach in the context of geotechnical engineering can be found in Baecher & Chirstian (2003), Baecher (2017) and Zhang (2017).

The Bayesian approach of statistical analysis refers to the method of statistical inference based on the Bayes rule, which describes a construct using the concept of conditional probability. The rule takes its name from the English mathematician Thomas Bayes who described it in his work published in 1763, two years after his death (Bayes 1763). The rule expresses the probability of a hypothesis given the data (*p[k|d]*) as a function of the probability of the data given the hypothesis (*p[d|h]*), the probability of the hypothesis (*p[h]*) and the probability of the data (*p[d]*).

The general form of the Bayes equation is:

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)} \tag{2}$$

which can also be interpreted in the following manner (Kruschke 2015):

$$posterior = \frac{likelihood \times prior}{evidence} \tag{3}$$

The Bayes rule is used to update the knowledge of a hypothesis (i.e. a model or a set of parameters) from observations represented by the data, and from the available prior knowledge on the hypothesis (i.e. subjective information or older data sets). The following sections present a brief description of the four components of the Bayes rule shown in Equation 3.

2.2.1 The posterior distribution

The 'posterior' is a probability distribution that reflects the uncertainty of the hypothesis examined (e.g. the set of parameters of a regression model) after taking into account the relevant data and prior knowledge on the hypothesis. The posterior is the answer sought by the analyst, reflecting the balance between the knowledge provided by the data and prior components. For this reason, the posterior is useful to gauge the sufficiency of data, as a strong dataset outbalances the effect of the prior. The uncertainty of the parameters inferred with a Bayesian analysis is reflected in the spread of the respective posterior distributions, which is measured with the HDI calculated for a particular level of confidence. The 95% HDI is commonly used to describe the uncertainty of parameters inferred from data and corresponds to the range of possible parameter values with a 95% credibility. The HDI has a similar purpose to the CI within the frequentist approach but its interpretation is simpler and more intuitive.

2.2.2 The likelihood function

The 'likelihood' function defines the probability of obtaining the observations included in the dataset given the hypothesis under examination (e.g. the set of parameters of a regression model). The likelihood is the answer given by classical statistical methods and reflects the likelihood of the hypothesis (i.e. the set of parameters) for that particular dataset.

Figure 1 shows an example extracted from Kruschke (2015) of the calculation of the likelihood of parameters of a normal distribution for a dataset of three points, d = [85, 100, 115]. In this case, the data points represent a variable x, which is assumed to follow a normal distribution with mean μ and standard deviation σ, hence:

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

The likelihood of a particular set of parameters [μ, σ] for a dataset of three points d = [x_1, x_2, x_3] corresponds to the product of the three probabilities of the data points as expressed by the likelihood function:

$$p(d|\mu,\sigma) = \prod_{i=1}^{3} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{5}$$

Therefore, the likelihood of an arbitrary chosen normal distribution with parameters μ = 87.8 and σ = 18.4, represented in blue in Figure 1, is 2.70E-06 for the dataset of three points shown in this figure. It is possible to verify that the likelihood of the calculated mean and standard deviation of the data points (μ = 100, σ = 12.2), represented in grey in Figure 1, corresponds to the maximum possible likelihood value, which is a known attribute of these parameters from classical statistics.

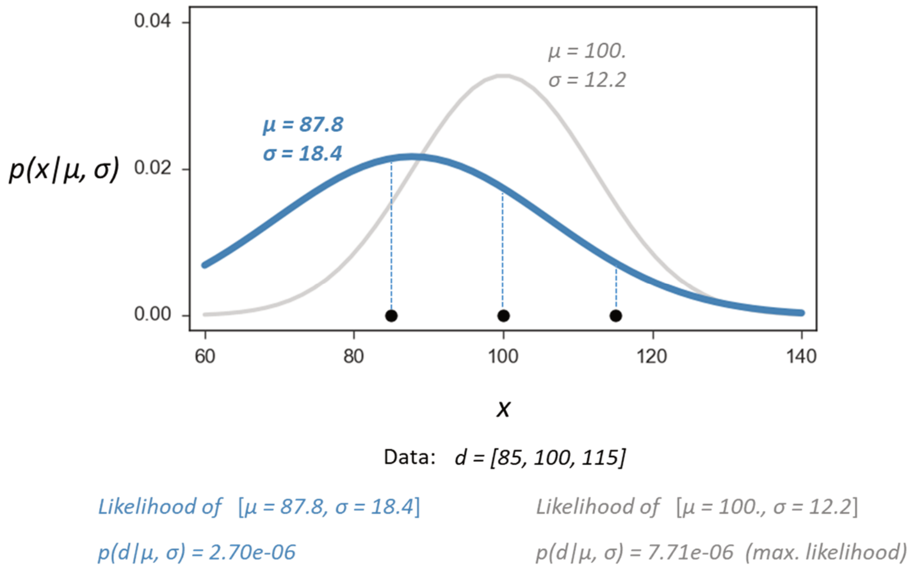


Figure 1 Example of calculation of the likelihood of the mean (μ) and standard deviation (σ) of a normal distribution assumed to represent the variability of a dataset of three points

2.2.3 The prior distribution

The 'prior' represents the initial knowledge on the hypothesis, and it can be informative or vague. Informative priors can be any type of distribution that represents adequately the existing knowledge of the model or parameter examined. Before the widespread availability of numerical methods to sample the posterior distributions, the selection of informative priors was based on their affinity with the likelihood function to facilitate the analytical calculation of the posterior. These priors are known as conjugate distributions.

The non-informative priors used to express ignorance about a parameter value are based on the range of the parameter domain, with the uniform distribution among the more commonly used for this purpose. However, there are situations where a different type of distribution may be required to contain the available information without the risk of over-constraining the results of the analysis. In these cases, the definition of the prior distribution could be based on the principle of maximum entropy, also known as the principle of minimum prejudice, developed by ET Jaynes in 1957 (Jaynes 1957). In this case, entropy refers to disorder or randomness in the information and is similar to the concept of entropy in physical systems. Table 1 shows a list of common maximum entropy probability distributions for various constraints, adapted from Harr (1987). The uniform distribution is a commonly used prior distribution and represents a non-informative condition in which only the limits of the parameter are known.

Table 1 Maximum entropy probability distributions

Constraints	Maximum entropy probability distribution
a ≤ x ≤ b	Uniform
x ≥ 0, mean known	Exponential
-∞ ≤ x ≤ +∞, mean and standard deviation known	Normal
a ≤ x ≤ b, mean and standard deviation known	Beta
0 ≤ x ≤ n, mean occurrence rate of independent events known	Poisson

The selection of the prior is an important step in a Bayesian analysis. The prior could add valuable available information to the posterior if selected adequately, or it could bias the results if it over-constrains the data. Figure 2 shows a conceptual representation of the influence of the prior on the posterior. The left column plots illustrate the situation of a vague prior having no influence on the posterior regardless of the size of the dataset. The middle column plots show the strong influence of an informative prior on the posterior when the dataset is small. The right column plots represent the case of an informative prior out weighted by the strong influence of a large dataset. The selection of inappropriate priors could result in over-constrained posterior distributions, in particular when data is scarce.

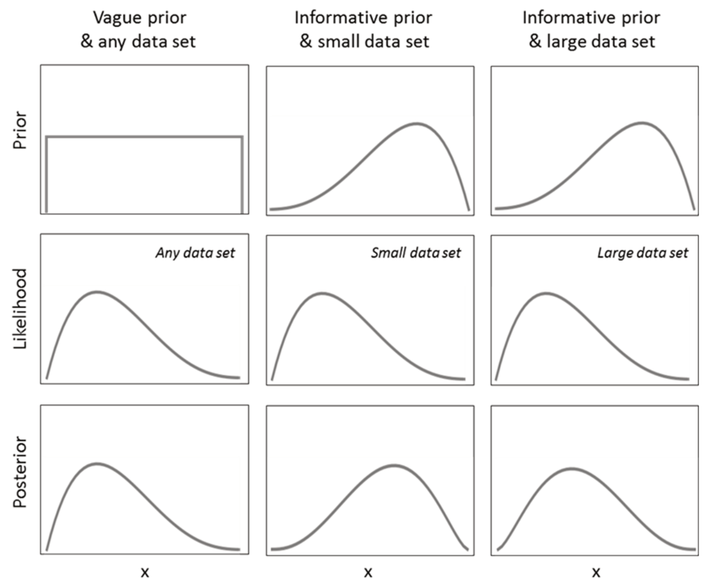


Figure 2 Conceptual representation of the influence of vague and informative priors on the posteriors depending on the size of the dataset

2.2.4 The evidence function

The 'evidence' part in the denominator of the Bayes equation (Equation 2) is normally treated as a normalisation factor so that the posterior integrates to one. It is calculated as the integral of the numerator over the whole parameter space. The posterior distribution does not need to be normalised when the purpose of the Bayesian analysis is the inference of parameters and the posterior is evaluated using the MCMC method. In this case, the calculation of the typically complex integral in the denominator of the Bayes equation can be skipped. However, the denominator is required when the objective of the analysis is the comparison of two alternative models, which is done through the calculation of the Bayes factor that relates the posteriors of the two models.

3 Assessment of sufficiency of data

There are various perspectives of the problem of assessing the sufficiency of data to ensure the level of confidence required in the design at a particular project stage. The following are the two fundamental perspectives on the problem.

3.1 The frequentist perspective

3.1.1 Interpretation of the confidence interval

It is intuitively known that the larger the number of data values (n) used to define a rock parameter such as the UCS, the larger the confidence of the estimate of the parameter and consequently the narrower the interval of possible values of the parameter. The measurement of the confidence of the estimate with the frequentist approach is based on the construction of the CI, which is said to have a particular probability (confidence level) of including the true mean of the property. However, this interpretation is misleading because the CI is specific to a dataset and its confidence level only has meaning in repeated sampling. This means that once the CI is calculated for a particular dataset there are only two possibilities; the CI either contains the true population parameter or doesn't. The confidence level refers in this case to the probability that CIs constructed for similar datasets from additional sampling contain the true but unknown population parameter.

The correct interpretation of the CI is explained with the example shown in Figure 3 based on a Monte Carlo analysis consisting in randomly generating 100 datasets of 10 UCS samples assuming that the population parameters are known (adapted from Contreras et al. 2018). In this case, the population parameters are represented by a mean and standard deviation of UCS of 206.4 and 45 MPa, respectively. These values correspond to the Norite set used by Gill et al. (2005) in their simulations and define a true coefficient of variation of 21.8% for this rock type. The dots in Figure 3 correspond to the means of the datasets of 10 values and the vertical grey bars represent the respective 95% CIs. The five CIs highlighted in red are the ones that do not contain the true mean, therefore 95% of the 100 datasets generated do contain the mean as illustrated in Figure 3. This example illustrates the true interpretation of the 95% confidence level of the CI. However, in a real case situation the true mean is unknown and the analyst cannot know the actual situation of the CI for the particular dataset under analysis relative to the true mean.

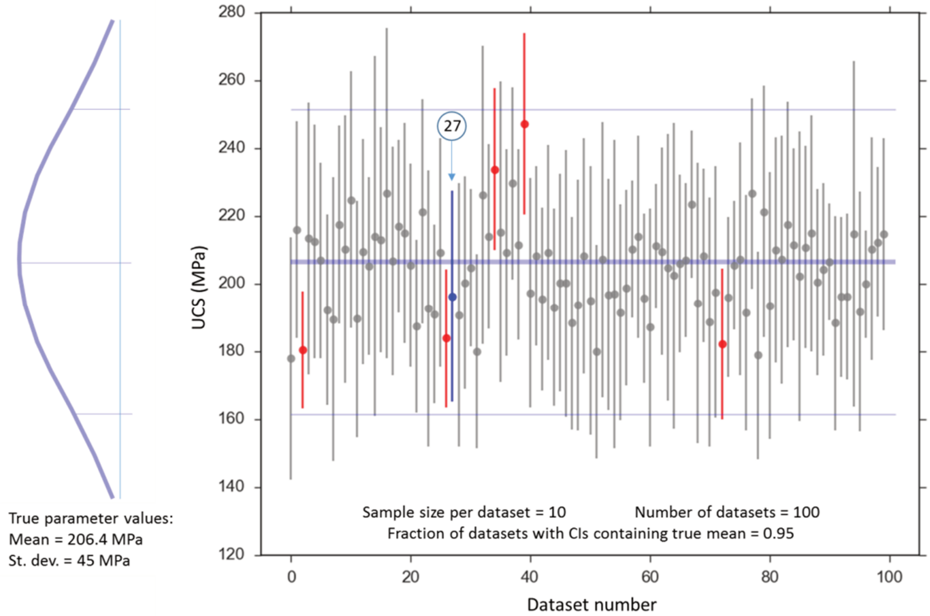


Figure 3 Interpretation of the CI in the frequentist approach based on the random generation of 100 datasets of 10 UCS values from a normal distribution with a mean of 206.4 MPa and a standard deviation of 45 MPa representing the true UCS population

3.1.2 Minimum sample size from the small sampling theory

Gill et al. (2005) describe a methodology to estimate the minimum number of test specimens required to determine rock properties with a laboratory testing program. The methodology is based on calculating the minimum number of tests required to reach a prescribed 'width' of the CI of the mean for a specified level of confidence. The CI is calculated in accordance with the concept of the small sampling theory using Equation 1. The width of the CI is determined by the precision index (*p*), which is defined as the ratio between the upper and lower bound values of the CI. The implication of this procedure is that in order to find an answer to the question on 'what is the appropriate number of tests', in addition to the level of confidence it is also required to define 'what is the appropriate precision index' that needs to be considered at different project stages. Gill et al. (2005) provide general guidelines in terms of the precision index recommended for different situations, indicating that $p \leq 1.5$ for temporary mine openings, $p \leq 1.35$ for permanent mine openings and $p \leq 1.2$ for important excavations requiring public safety, with the level of confidence at 95% in all cases.

The number of data values, the 'width' of the CI, and the confidence level are three elements interrelated through the relationships from classical statistics for construction of the CIs. The prescription of two elements is required to define the third. The method of Gill et al. (2005) consists of prescribing the level of confidence (typically 95%) and the width of the CI (recommended p) to estimate the sample size required to determine the parameter value with the required confidence. However, it is difficult to estimate the optimum size of the dataset with certainty because of the changes in the coefficient of variation (CV) of the sample when new data points are added, in particular for small datasets. Figure 4 shows the relationship between *p*, *n* and *CV* for a 95% confidence level and illustrates the aspects affecting the width of the CI.

Gill et al. (2005) proposed an iterative procedure to find the optimum sample size for a particular dataset. The method is intended to get consistency between *n*, *p* and *CV* for the optimised sample size. However, a notable limitation of the method is that the calculated optimum number of data points is only valid for that particular set of samples and the result could be different with a different dataset. This is a drawback derived from the conceptual basis of the frequentist approach for inference of parameters and it is analogous to the drawback of the CI to measure the confidence of the true population parameter as explained earlier.

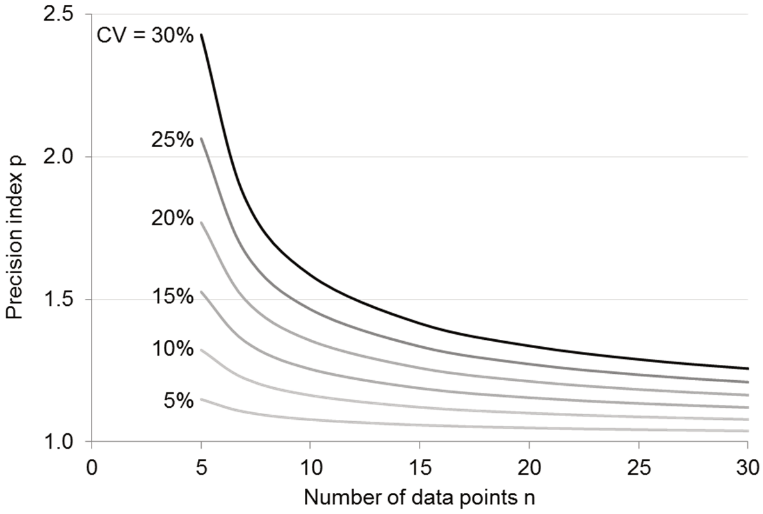
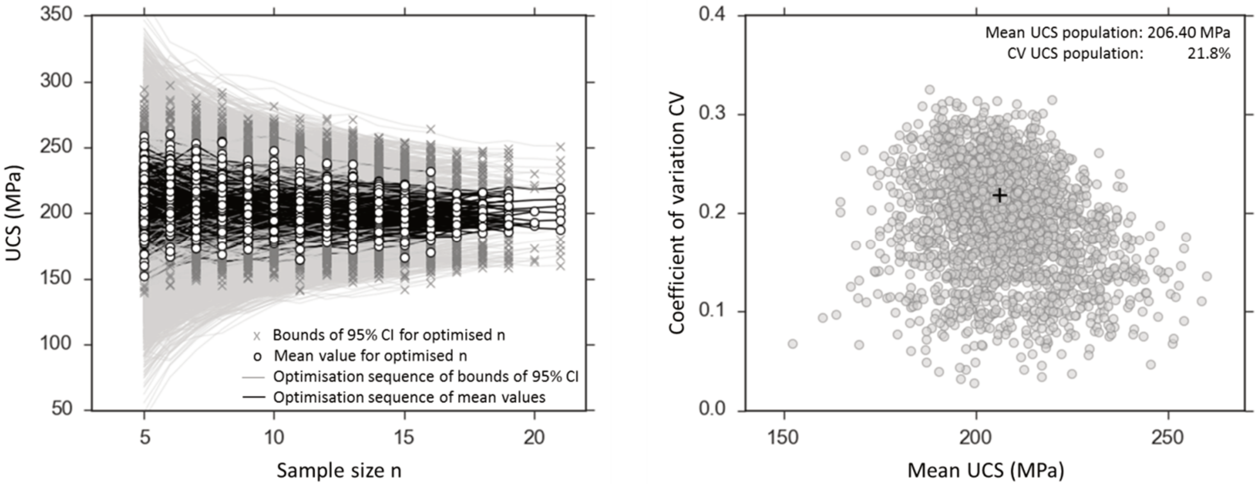


Figure 4 Influence of sample size (*n*) and sample variability (*CV*) on the width of the 95% CI (*p* or the ratio of the upper and lower bounds of the CI)

The uncertainty of the optimum sample size based on the frequentist approach is demonstrated with a Monte Carlo analysis simulating 2,000 datasets sampled from the Norite population parameters used in the example of Figure 3. Each dataset starts with five data points and sampled values are added to the dataset until the optimum sample size is reached according to the method of Gill et al. (2005), considering a confidence level of 95% and a target p of 1.35. The results of this analysis are shown in Figure 5. The dots in Figure 5a represent the means of the optimised datasets, the dark lines correspond to the mean sequences from the initial dataset of five points up to the optimised dataset, the crosses mark the upper and lower bounds of the optimised CIs and the grey lines define the sequences of the CI bounds from the initial to the optimised dataset sizes. Figure 5b presents the scatter plot of the mean UCS versus the CV for the optimised datasets, showing the spread around the true population parameters represented by the dark cross in the middle.



(a) (b)

Figure 5 Uncertainty in the calculation of the optimum sample size with the method of Gill et al. (2005) for a target precision index p = 1.35 based on a Monte Carlo analysis with 2,000 sampling sequences. (a) Progression of bounds of the 95% CI and mean values of UCS from a minimum sample size of five data points up to the optimised sample size. (b) Scatter plot of coefficient of variation CV versus mean UCS for the optimised sample sizes

The results in Figure 5a show the effect of the variability of the datasets on its calculated optimum size. This aspect is illustrated more clearly in Figure 6a that shows the relationship between the CV of the optimised datasets and the optimum sample size (n_{opt}). The datasets with low CVs define narrow CIs, causing that the target p is reached with few data points, and in some cases, the initial five values are already exceeding the $p \leq 1.35$ condition. Figure 6b shows that the optimised datasets with few numbers of data points exceed the target p by an ample margin because the width of the CI is more sensitive in this region to the change of the number of data points as shown in Figure 4. The average optimised sample size for the example presented in Figures 5 and 6 is 10 data points, but the variability is high with optimum sample sizes varying between 5 and 21. However, in a real case situation, the number of data points is limited and the estimation of the optimum sample size itself has a big uncertainty.

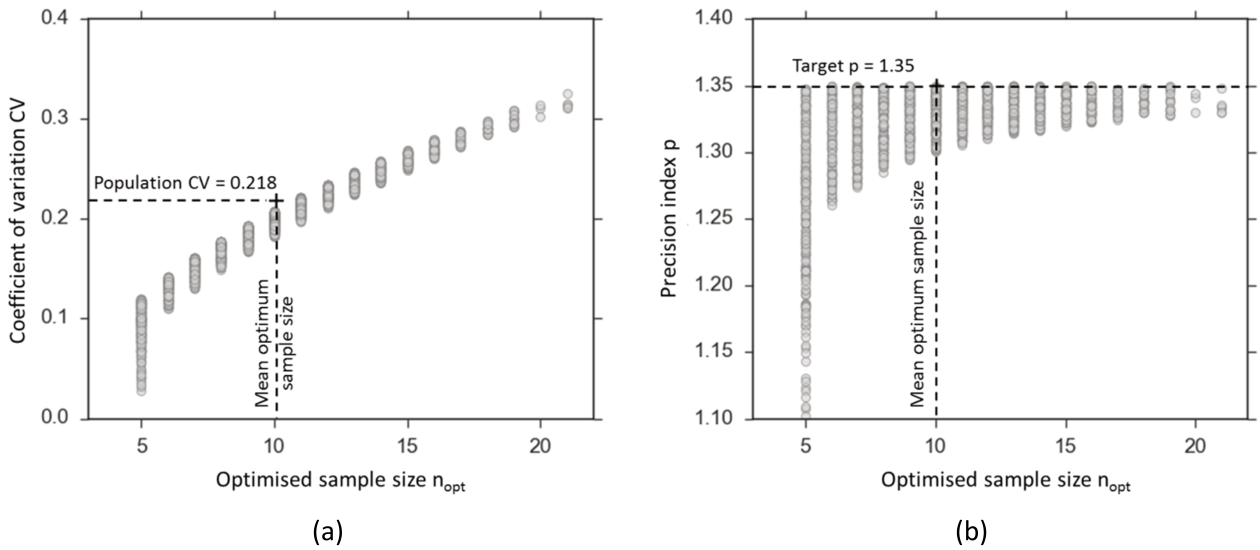


Figure 6 Effect of the variability of the sample on the determination of the optimised sample size for a target precision index p = 1.35. (a) Graph showing the correlation between the coefficient of variation CV of the sample and the optimised sample size n_{opt}; (b) Graph showing the effect of n_{opt} on the actual *p* reached with the method of analysis

3.1.3 Minimum sampling size from other methods

Gill et al. (2005) include a detailed discussion on alternative methods and heuristic rules from different organisations and authors on the minimum number of data points required to determine particular rock properties from laboratory testing. Some of these methods are based on variations to the calculation of the CI but still within the framework of the frequentist approach of statistical analysis and others correspond to practical recommendations based on specific experience and judgement in the area of application of the guidelines, but still with background support from frequentist concepts.

The method based on a frequentist statistical framework relay on data being the result of a random sampling process and therefore, predictions on the sufficiency of data based on a specific dataset are inherently uncertain requiring multiple repeated sampling for the confidence level to be meaningful. The Bayesian approach provides the conceptual basis for a clearer, more intuitive assessment of data sufficiency from limited data.

3.2 The Bayesian perspective

3.2.1 The highest density interval (HDI) to measure uncertainty

The method discussed in the previous section to relate the uncertainty of the parameters with the number of data points is based on the concept of the CI in repeated sampling. In general, the CI is mistakenly used to quantify the uncertainty of parameters such as the mean value of a rock property, when in fact the CI really measures the uncertainty of the data supporting the parameter estimate. The confusion is the result of the intuitive interpretation by the analyst of data as fixed and parameters as random entities, which is inconsistent with the assumptions of the approach. However, this intuitive interpretation of data and parameters is consistent with the assumptions of the Bayesian approach, which suits better the interest of the analyst as a result.

Figure 7 shows the contrast between the 95% CI from the frequentist (classical) approach and the 95% HDI from the Bayesian approach for the estimation of the mean UCS with dataset 27 in Figure 3 (highlighted in blue). The histogram in the frequentist analysis represents the distribution of the 10 data points in the dataset and the result corresponds to a point estimate of the mean and the measure of the error in the estimation reflected by the 95% CI. The Bayesian result corresponds to the posterior probability distribution of mean values derived from the same dataset of 10 values and assuming a non-informative prior represented by a uniform distribution between 50 and 350 MPa. The 95% highest concentration of values in the posterior distribution is a good representation of the uncertainty of the mean UCS and defines the 95% HDI. This interval seems to coincide with the frequentist 95% CI; however, the meaning of the two results is different. The 95% HDI defines the range of the mean UCS with a 95% credibility considering the valid range specified with the prior distribution and the 10 data points in dataset 27. In contrast, the 95% CI marks the interval based on the 10 data points in dataset 27 that may contain the unknown true mean UCS with a 95% confidence; however, this confidence is meaningful only when the CIs of similar datasets from future sampling are examined. In the Bayesian approach, the mean is variable and it could be updated as new data become available. In the frequentist approach, the mean is fixed and the random sampling represented by data is used to make inferences of its possible value.

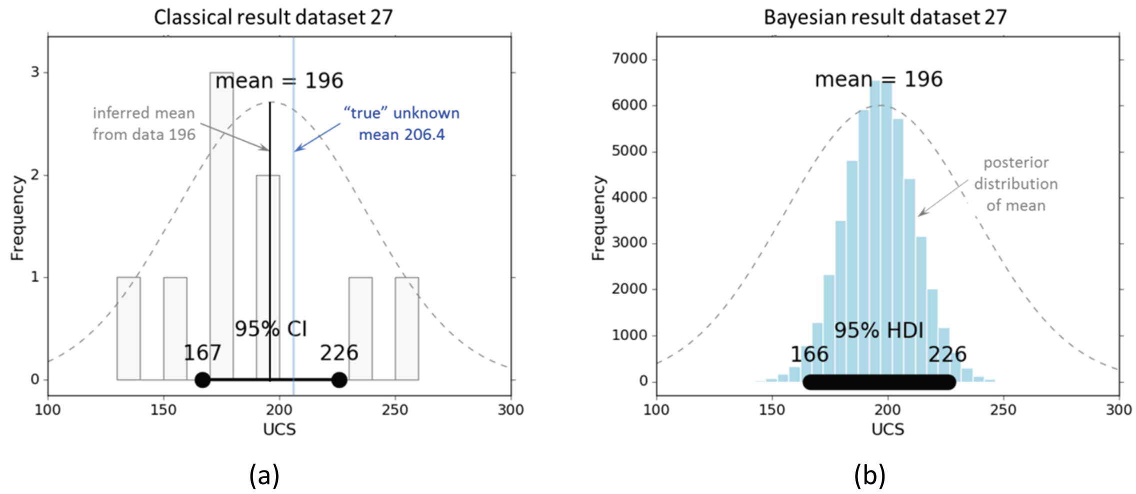
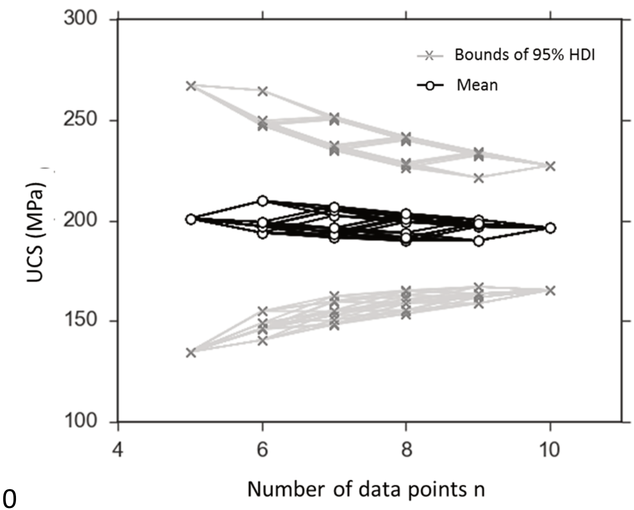


Figure 7 Comparison of results of inference of the mean UCS with 10 data points corresponding to dataset 27 in Figure 3. (a) Point estimate of mean and CI from frequentist (classical) approach; and (b) Posterior probability distribution of mean from Bayesian approach

3.2.2 Assessment of sufficiency of data from updating the posterior

The proposed method to assess the sufficiency of data with the Bayesian approach is based on the assessment of the variation of the mean and width of the 95% HDI of the posterior distribution of the parameter investigated, as additional information is available. The posterior distribution reflects the balance between the available knowledge included in the prior distribution and data. Therefore, the changes in the posterior distribution with the addition of new data points reflect the strength of the dataset to outbalance or support the prior information. The posterior in Figure 7b considered a non-informative prior (uniform distribution from 50 to 350 MPa) and, therefore, this example is a case of data reinforcing the vague available knowledge.

Dataset 27 was selected arbitrarily from those represented in Figure 7 to illustrate the change of the posterior from a base dataset of five values up to the 10 available data points in the dataset. Figure 8 lists the values in this dataset and shows the variation of the mean and width of the 95% HDI from this analysis. In this case, the first five data points represent the base dataset and the remaining five data points are added according to the 120 different possible sequences of these values. The variation of the characteristics of the posterior relative to the previous data stage offers a guideline for assessing the sufficiency of data. For example, the curves in Figure 8 suggest that the mean and width of the HDI tend to a stable condition from eight data points, confirming that the sample size of 10 data points is adequate to define the mean UCS with the prescribed level of confidence. It may be argued that a similar result would be obtained by using the CI from the frequentist approach. However, the difficulty, in that case, is to assign a meaningful interpretation to the result considering that the confidence level has meaning only for repeated sampling as discussed in Section 3.1.



0

Dataset 27 (Figure 3)

No.	UCS (MPa)	
1	239.04	Base dataset
2	176.00	
3	268.01	
4	132.05	
5	190.22	
6	253.45	Values used for random selection
7	177.02	
8	158.38	
9	178.86	
10	190.59	

Figure 8 Variation of the mean and the width of the 95% HDI of the posterior distribution of UCS from a Bayesian analysis with the number of data points for dataset 27 in Figure 3. The curves correspond to eight random arrangements of data points no. 6 to 10

An example of the use of the progression of the posterior distribution to assess the sufficiency of data is described by Contreras & Brown (2019) with reference to the Bayesian inference of the intact rock strength parameters σ_{ci} and m_i of the Hoek–Brown criterion. The example considers three stages with increased levels of data to show the relationship between data quantity and the uncertainty of the estimation. The analysis includes three data sets with 10, 18 and 23 data points consisting of UCS and triaxial compression strength (TCS) tests as detailed in the results of the fitting analysis reproduced in Figure 9.

The graphs in Figure 9a show the data, the mean fitted envelopes and the bands defining the uncertainty of the estimation corresponding to the 95% HDI of the parameter posteriors for the three levels of data. The scatter plots in Figure 9b show the sampled parameter values from the posteriors reflecting the reduction in the uncertainty of the estimation with the increased levels of data. The scatter plots include the contours marking the 68% and 95% concentration of points. The parameter values within the 95% contours in Figure 9b define the width of the envelope bands in Figure 9a. In general, these results show a good correlation between data quantity and the uncertainty of the strength parameters defining the envelopes. For example, the increase of the number of TCS results used for the regression analysis has a direct effect on the reduction of the uncertainty of the envelope in the high confining stress region. In the low confining stress region, the intercept of the envelope associated with σ_{ci} is well defined with relatively few UCS data points.

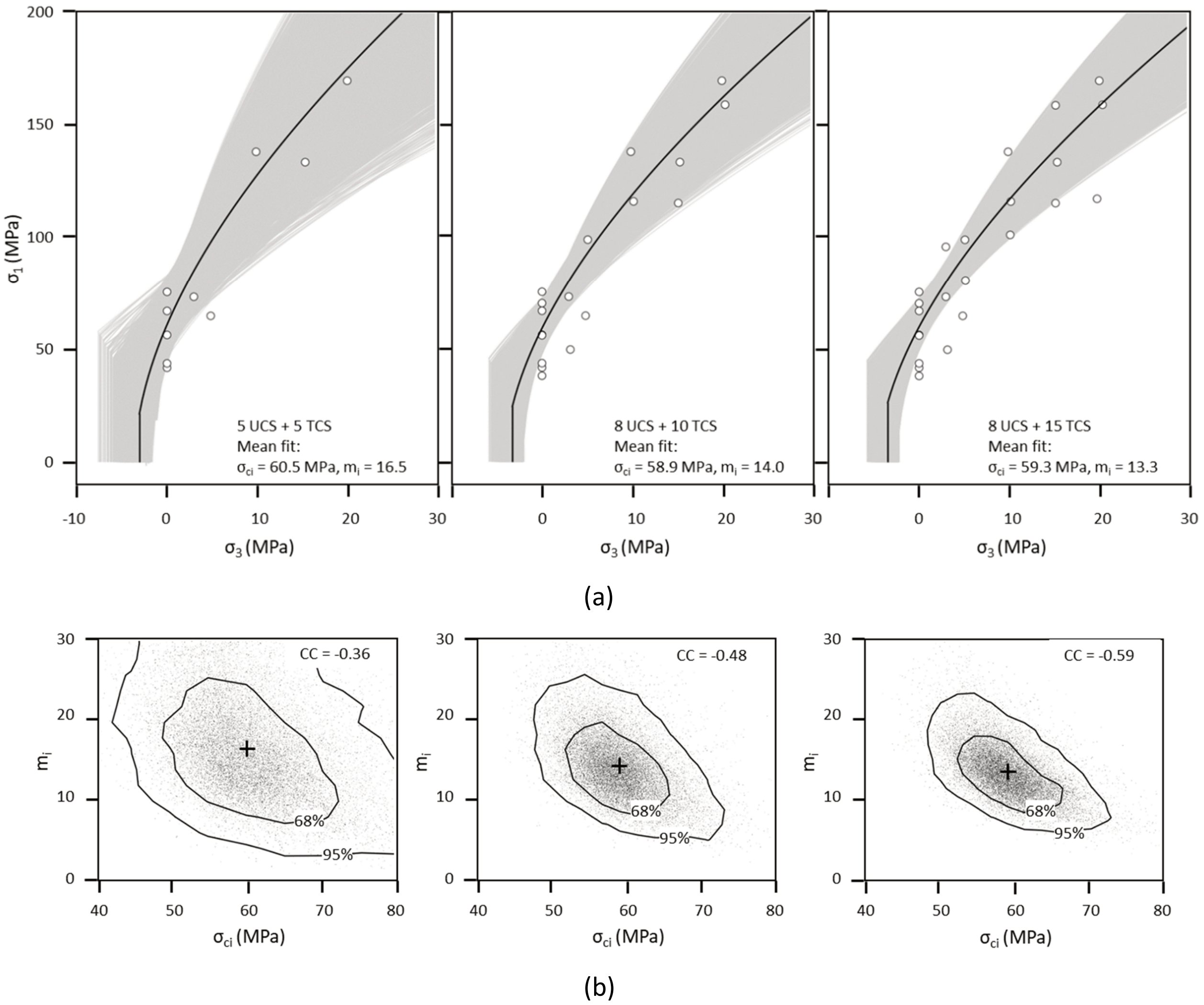


Figure 9 Relationship between data quantity and uncertainty of the estimation. (a) Mean fitted envelopes with bands including the 95 percentile of sampled parameter values for three levels of data; (b) and Scatter plots of m_i versus σ_{ci} from the Bayesian regression analysis with 68 and 95 percentile contours and coefficients of correlation CC (Contreras & Brown 2019)

Figure 10 shows a comparison of the mean fitted envelopes from the analysis with the three levels of data. The envelopes are close in the low confining stress region and the differences are associated with the number of TCS data points used in the analysis. These results suggest that at least 10 TCS data points are required in this particular case to define a reliable mean envelope. A sufficient number of TCS results is required to outbalance the effect of the vague priors of σ_{ci} and m_i as seems to be the case for the second and third stages with 10 and 15 TCS results, respectively.

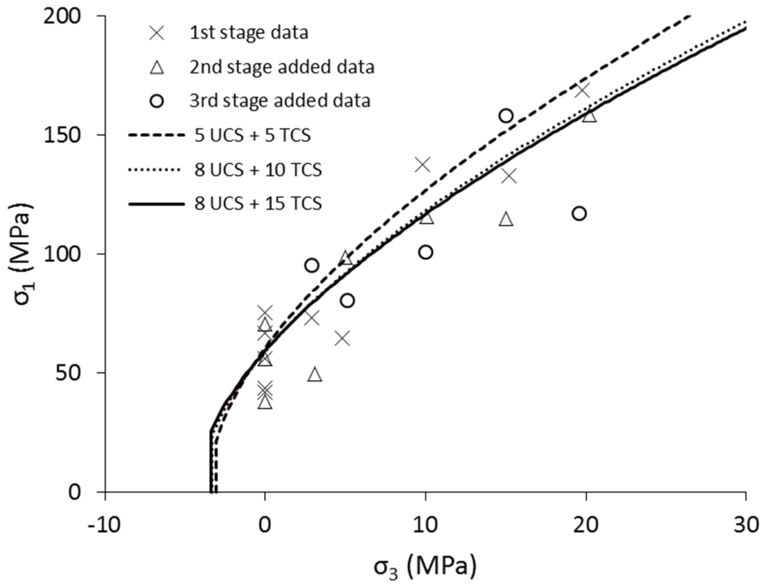


Figure 10 Mean fitted envelopes for three stages with increased levels of data (Contreras & Brown 2019)

The graphs in Figure 11 show the changes in the mean value and the width of the 95% HDI of the inferred parameters σ_{ci} and m_i for an increase in the number of data points from five to 23. The data points include UCS and TCS test results, which have a predominant influence on σ_{ci} and m_i, respectively. These graphs show that the mean values and the variability of the parameters (width of 95% HDI) tend to a stable situation after a certain number of data points, which is a pattern useful for the assessment of the sufficiency of data.

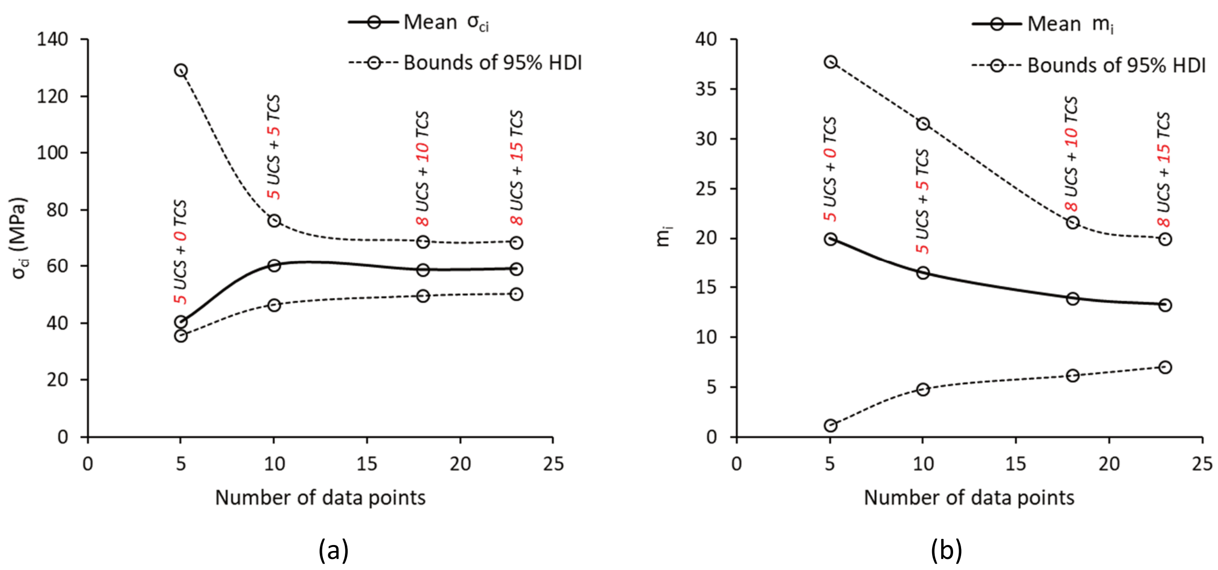


Figure 11 Relationship between the number of data points and the variability of the intact Hoek–Brown rock strength parameters from the example presented in Section 6.2.2. (a) Parameter σ_{ci}; (b) Parameter m_i

The graphs in Figure 11 provide a qualitative reference to define the sufficiency of laboratory testing data for the characterisation of the intact rock strength. However, they could also be useful to derive specific criteria to assess the number of test results required for a particular project stage. For example, the percentages of change of the mean and width of HDI could serve as a measure of the degree of convergence of the estimates. Table 2 shows the percentages of change relative to the previous data stage, normalised to the number of added data points, for the intact rock strength parameters inferred according to the stages depicted in Figure 11.

Table 2 Percentages of change of inferred parameters with the number of data points

Data stage	No. data points	Type of data	% change σ_{ci}		% change m_i	
			mean	HDI width	mean	HDI width
1	5	5UCS+0TCS				
2	10	5UCS+5TCS	6.6%	-21%	-4.2%	-12%
3	18	8UCS+10TCS	-0.3%	-2%	-2.3%	-10%
4	23	8UCS+15TCS	0.1%	-0%	-1.0%	-4%

The Bayesian approach shows a clear relationship between parameter variability and data quantity, manifested through the convergence of the parameter values and the associated width of the posterior HDIs. These characteristics provide a powerful tool to assess the sufficiency of data in an intuitive way. This degree of convergence with the increase of the size of the datasets could be quantified as shown in Table 2, and these measures could be used for the definition of criteria to assess the sufficiency of data, although this objective is not part of this paper.

4 Conclusion

The Bayesian approach of statistical analysis is more suitable for the quantification of the geotechnical uncertainty in slope design as compared with the classical approach (frequentist). The two approaches are based on different definitions of probability and different sets of assumptions that suit different types of uncertainty. The scarcity of data is a common occurrence in the slope design process and the corresponding knowledge uncertainty (epistemic) derived from this situation is better treated with Bayesian methods. In contrast, classical methods are meant to treat aleatory uncertainty (natural variation), which implies the availability of abundant information for its characterisation. The methodology also shows the relationship between data quantity and the uncertainty of the inferred parameters in a way that can be used to assess the sufficiency of data.

The methods commonly used to define the adequate number of data points required to estimate design parameters of rock with a prescribed level of confidence are based on the concept of CI from the frequentist (classical) approach of statistical analysis. However, there are various drawbacks of this approach including that the width of the CI needs to be prescribed, which adds subjectivity to the calculation. Moreover, the intuitive interpretation normally given to the CI is inconsistent with the conceptual basis of the frequentist approach and rather corresponds to the interpretation of the HDI from the Bayesian approach. In the frequentist framework, data is considered to be the result of a random sampling process and, therefore, predictions on the sufficiency of data based on a specific dataset are inherently uncertain requiring multiple repeated sampling for the confidence level to be meaningful. The Bayesian approach provides the conceptual basis for a clearer, more intuitive assessment of data sufficiency from limited data.

The results of the Bayesian analysis enable a rational assessment of the sufficiency of data at different stages of project development. The results reflect the balance between prior information and data; therefore, if data is weak, the prior knowledge dominates the result. As more data is collected the Bayesian results move toward stable values that are unaffected by the prior component. The prior serves in this case to test the strength of data and this behaviour provides the analyst with a good reference to judge the adequacy of the dataset.

The Bayesian approach shows a clear relationship between parameter variability and data quantity, manifested through the convergence of the parameter values and the associated width of the posterior HDIs. The convergence of the mean and spread of the posterior distributions could be measured for the purpose of defining criteria to assess the sufficiency of data. However, the main purpose of this paper is to show the benefits of the Bayesian approach to reflect the relationship between data quantity and uncertainty of the estimation in an intuitive way and with a clear interpretation consistent with the conceptual basis of the approach.

Acknowledgement

The first author was supported by the Large Open Pit II project (LOP II) through contract no. 019799 with the Geotechnical Research Centre (GEC) of the University of Queensland and by SRK Consulting, South Africa. Their support is greatly appreciated.

References

Baecher, GB, 2017, 'Bayesian thinking in geotechnics', *Proceedings of Georisk 2017,* American Society of Civil Engineers, Reston.

Baecher, GB & Christian, JT 2003, *Reliability and Statistics in Geotechnical Engineering*, Wiley, Chichester.

Bayes, T 1763, 'An essay towards solving a problem in the doctrine of chance', *Philosophical Transactions*, vol. 53, pp. 370–418.

Contreras, LF & Brown, ET 2019, 'Slope reliability and back analysis of failure with geotechnical parameters estimated using Bayesian inference', *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 11, no. 3, pp. 628–643.

Contreras, LF, Brown, ET & Ruest, M 2018, 'Bayesian data analysis to quantify the uncertainty of intact rock strength', *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 10, no. 1, pp. 11–31.

Gill, DE, Corthesy, R & Leite, MH 2005, 'Determining the minimal number of specimens for laboratory testing of rock properties', *Engineering Geology*, vol. 78, pp. 29–51.

Harr, ME 1987, *Reliability-based Design in Civil Engineering*, Dover Publications Inc., Mineola.

Jaynes, ET 1957, 'Information theory and statistical mechanics', *The Physical Review*, vol. 106, no. 4, pp. 620–630.

Kruschke, JK 2015, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS and STAN*, Academic Press, Cambridge.

Read, J 2009, 'Data uncertainty', in J Read & P Stacey (eds), *Guidelines for Open Pit Slope Design*, CSIRO Publishing, Collingwood, pp. 211–218.

VanderPlas, J 2014, *Frequentism and Bayesianism: A Practical Introduction*, blog post, viewed 20 January 2015, http://jakevdp.github.io/blog/2014/03/11/frequentism-and-bayesianism-a-practical-intro/

Zhang, J 2017, *Bayesian Method: A Natural Tool for Processing Geotechnical Information*, joint TC205/TC304 Working Group on "Discussion of statistical/reliability methods for Eurocodes", Final Report, International Society for Soil Mechanics and Geotechnical Engineering, London.