

Article

Not peer-reviewed version

Rock Engineering Knowledge and Radical Uncertainty: From Empirical Methods to Professional Practice

[Davide Elmo](#)^{*} and Samantha Kenzie Adams

Posted Date: 14 November 2025

doi: 10.20944/preprints202511.1089.v1

Keywords: GSI; Hoek-Brown criterion; uncertainty; professional practice; synthetic rock mass modeling; calibration versus validation; rock engineering knowledge



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Rock Engineering Knowledge and Radical Uncertainty: From Empirical Methods to Professional Practice

Davide Elmo ^{1,*}, Samantha Kenzie Adams ²

¹ University of British Columbia, NBK Institute of Mining Engineering, Vancouver, Canada;

² University of Toronto, Faculty of Applied Science & Engineering, Toronto, Canada;

* Correspondence: davide.elmo@ubc.ca

Abstract

This paper examines the empirical foundations of rock engineering. We have adopted a narrative style throughout, as the philosophical nature of the questions raised in the paper is better served by discourse than conventional technical structure. In the first part of the paper, we revisit the Hoek-Brown failure criterion using synthetic rock mass models, validating the general framework while revealing critical distinctions between data-driven parameter emergence and classification-dependent derivation. The analysis reveals fundamental challenges, including the conflation of calibration with validation, the transformation of qualitative geological assessments into seemingly quantitative parameters, and the false similarity problem, where different parameter combinations yield equivalent failure envelopes. In its second part, the paper explores broader implications for professional practice, revealing a significant gap between methodological validation and professional acceptance. What satisfies the professional standard of "balance of probabilities" for establishing reasonable practice may fall short of the scientific standard of "evidence beyond reasonable doubt" required to claim genuine predictive validity. We propose "epistemological integrity" as a framework for responsible practice: acknowledging the false similarity problem, communicating uncertainty transparently, applying validation standards consistently, and aligning professional claims with actual knowledge limitations rather than projecting false precision through computational sophistication.

Keywords: GSI; Hoek-Brown criterion; uncertainty; professional practice; Synthetic rock mass modeling; Calibration versus validation; Rock engineering knowledge

1. Introduction

Rock engineering practice operates within a fundamental epistemological contradiction. While our analytical frameworks require numerical inputs, the geological characterizations underpinning these numbers derive from qualitative assessment and empirical assumptions that not classify as genuine quantitative measurements [1]. This led to the conversion of subjective empirical observations into seemingly objective numerical parameters used in our design calculations. The contradiction deepens when we apply measurement terminology as "accurate", "precise" and "reliable" to these qualitative assessments.

The increasing use of artificial intelligence (AI) and its subset of machine learning (ML) in rock engineering intensifies these epistemological challenges. As the profession increasingly adopts automated systems, predictive algorithms, and AI-assisted design tools, we risk encoding and amplifying the contradictions and knowledge limitations already embedded in design methods commonly accepted by researchers and practitioners [2, 3, 4]. For instance, when datasets used for

training purposes consist of subjectively derived parameters, AI systems inherit our epistemic uncertainties and systematic biases. Furthermore, due to their inherent nature, many AI systems can mask the subjective foundations upon which their predictions are based, creating an illusion of objectivity that may be more dangerous than an explicit acknowledgment of uncertainty. More importantly, [3] noted that even when AI applications to rock engineering problems reference physical evidence, independent validation of results often proves impractical.

Professional challenges compound these concerns. As mining operations extend to greater depths and more complex geological environments, the knowledge gap between our empirical databases and actual operating conditions continues to widen. Meanwhile, our design methods are ill-equipped to address new uncertainties related to time-dependent rock mass damage and societal demands for environmental protection, which have broader and more complex consequences than traditional design methods can account for [5].

The inherent uncertainty of our geological/geotechnical assessments, the proliferation of AI tools trained on datasets derived from largely qualitative assessments and empirical observations, and professional responsibilities give rise to an epistemological three-body problem for which no general analytical solution exists within standard engineering frameworks. This recognition demonstrates that rock engineering knowledge is neither purely scientific nor purely empirical. Rather than viewing this as a limitation, we argue that acknowledging the boundaries of our knowledge may lead to more robust engineering practice that better serves public safety, environmental stewardship, and professional integrity in an era of increasing technological sophistication and societal scrutiny.

This paper draws inspiration from Flexner's 1939 essay [6], in which the author states that:

"Throughout the whole history of science, most of the really great discoveries which had ultimately proved to be beneficial to mankind had been made by men and women who were driven not by the desire to be useful but merely the desire to satisfy their curiosity."

This matters for rock engineering. Our industry and research establishment expect every research paper to deliver an immediate value proposition, promise near-term impact, and yield practical applications. However, by systematically directing resources only toward practical outcomes, we reveal the troubling assumption that either we have solved the fundamental problems and all we need are minor tweaks, or that we will never. Either way, we ensure we never escape our fundamental uncertainties. The problem is not that rock engineering remains empirical, but that in too many instances we have stopped being empirical, accepting historical correlations without the ongoing data collection that empiricism demands.

Within this context, the objectives of this paper are threefold: first, to examine the philosophical foundations of rock engineering knowledge and expose the epistemological challenges inherent in widely adopted design methods; second, to analyze how operational definitions, validation processes, and numerical modeling approaches may generate misleading precision rather than meaningful understanding; and third, to propose a framework for acknowledging and working within the boundaries of geological uncertainty rather than attempting to eliminate it through increasingly sophisticated analysis. More specifically, our examination addresses three critical research questions:

- What constitutes appropriate validation for rock engineering theories and models when direct experimental verification is impossible?
- How can rock engineering practice develop frameworks for decision-making under radical uncertainty that preserve safety while acknowledging the inherent limitations of our predictive capabilities?
- How do the epistemological limitations of rock engineering knowledge affect the validity and reliability of design tools assisted by AI systems, and what safeguards are necessary to prevent the amplification of cognitive biases in automated decision-making?

This paper is organized into two parts: the first revisits the empirical foundations of rock engineering, with a particular focus on the Hoek-Brown failure criterion [7], while the second addresses the broader implications of rock engineering knowledge for professional practice. We have

intentionally adopted a narrative style, as the nature of these questions is better served by discourse than by the conventional structure of technical papers.

2. The Empirical Structure of Rock Engineering Knowledge

Rock engineering inquiry should begin with experiments designed to answer specific questions. These experiments need not be confined to physical laboratory settings since the scales inherent in rock engineering problems necessitate large-scale field testing and synthetic experiments through numerical simulations. However, the fundamental question arises as to whether our observations represent knowledge or merely information. Observations divorced from conceptual frameworks remain mere data points in a database. Conversely, the validity of any theoretical framework depends entirely on the information we gather. This creates a circular dependency that requires both a quantitative structure from which to derive predictions and a theory of probability to assess the quality of these predictions.

Rock engineering's dependence on quantified qualitative assessments creates inherent vulnerability to theoretical bias [8]. As Popper [9] argued, scientific knowledge should not depend on personal judgment but be evaluated solely through mathematics and logic. The transition from laboratory observations to field-scale predictions introduces additional epistemological problems, as the validation of our models requires linking theoretical problems to observations through rules that connect not only to defining relations but also to conclusions logically deduced from those relations [10].

Proper validation transcends simple agreement or curve-fitting as it demands that underlying mechanisms be logically explainable and measurable. To understand calibration and validation challenges in rock engineering, we must first define these terms and recognise that their definitions vary across scientific and engineering disciplines. At the most fundamental level, calibration involves adjusting model parameters to match known observations or reference standards. It answers the question "Can we tune the model to reproduce what we have already observed?" Validation, conversely, tests whether a calibrated model can successfully predict independent observations not used during calibration, addressing the more demanding question "Does the model work for cases it has not seen before?"

This fundamental distinction, however, manifests differently across disciplines, reflecting varied priorities and practical constraints (Table 1). Measurement and instrumentation fields interpret these terms through the lens of equipment accuracy rather than predictive modelling. This interpretation emphasizes reproducibility and accuracy in relation to external references rather than predicting new scenarios. Computational modelling in engineering and physics introduces an additional layer of complexity by distinguishing verification from both calibration and validation. This framework acknowledges that computational accuracy (verification) is conceptually distinct from parameter fitting (calibration) and predictive capability (validation). In machine learning and data science, the process is formalized into three distinct stages: training (analogous to calibration), validation and testing. This three-tier structure reflects the field's emphasis on predictive capability and its ability to partition large datasets. The distinction between "validation set" and "test set" in machine learning represents a more granular approach than typically employed in traditional rock engineering practice.

Rock engineering practice occupies an uncomfortable middle ground among these disciplinary approaches. Unlike instrumentation calibration, our "measurements" of rock mass parameters involve qualitative geological assessments [11]. Unlike computational physics, we often conflate calibration with validation. Back-analysis illustrates this problem directly. Since back-analysis infers parameters from known failure conditions, it falls under the definition of calibration rather than validation, yet is often treated as the latter. More importantly, in rock engineering practice, validation does not operate in parallel; rather, it should be established independently for each site or condition, like a series circuit. And unlike machine learning, we rarely possess datasets large enough to partition into training, validation, and test subsets [4].

Table 1. Definition of Calibration and Validation across different disciplines.

Discipline	Calibration	Validation	Additional Layer
Measurement & Instrumentation	Adjusting an instrument against known standards (e.g., calibrating a scale with certified weights)	Confirming the instrument performs correctly across its operating range. The focus is more on equipment accuracy than predictive models	n/a
Computational Modelling	Adjusting parameters to match known behaviour	Comparing predictions to independent experimental/field data	Verification. It addresses whether equations are solved correctly
Machine Learning & Data Science	Fitting model parameters to training data (Training, analogous to calibration).	Testing on a validation set during model development to tune hyperparameters	Testing. Final evaluation on completely independent test data (closest to validation in other fields)

More significantly, rock engineering has developed what might be termed “validation through consensus”. As discussed by [8, 11], empirical methods in rock engineering practice are often accepted as validated through widespread professional acceptance rather than through systematic testing against independent data. This represents a departure from all the disciplinary frameworks described above. The question then becomes whether this disciplinary peculiarity reflects a pragmatic adaptation to rock engineering’s unique constraints (e.g., the difficulty of conducting large-scale field experiments, the scarcity of independent failure data, and the inability to derive predictive capability from case histories alone) or represents an epistemological compromise that undermines the validity claims our profession routinely makes.

This distinction between methodological validation and professional acceptance may carry significant implications for professional practice, as discussed later in Section 7. What could, in principle, satisfy the professional standard of “balance of probabilities” for establishing reasonable practice may fall short of the scientific standard of “evidence beyond a reasonable doubt” required to claim genuine predictive validity.

Successful calibration does not guarantee a physically realistic mechanistic representation [5]. Consider the scenarios in which fundamentally different conceptual models, continuum versus discontinuum, both successfully match observed behaviour through parameter adjustment. Such equivalence suggests that the models are fitting the data rather than capturing the underlying physics. One remedy involves continuous recalibration as new data emerge, which Elmo [5] terms “indirect validation.” However, two challenges persist: i) practitioners may bypass mechanistic validation even when site-specific data exist, and ii) the need for continuous calibration may indicate inadequate underlying relationships rather than merely incorrect parameters.

When considering computational modelling and data science, rigorous validation requires splitting available data into calibration and validation subsets before analysis begins. The calibration subset is used to develop the model, while the validation subset independently assesses its predictive performance. Applied to empirical rock engineering methods (e.g., the relationships used to derive the parameters m_b and s in the Hoek-Brown failure criterion, and pillar formulae), this would involve deriving correlations from one data subset and then testing them against a reserved data set spanning different rock types, fracture configurations, and scales. As [12] observes, to date, such a systematic validation process is absent in rock engineering practice.

A further fundamental limitation affects all validation efforts in rock engineering practice, which [5] identifies as a core constraint of empiricism. While models representing general physical trends can extend beyond calibration data, this principle applies specifically to mechanistic models

grounded in fundamental physics, rather than to empirical correlations. According to [5], validated empirical models remain valid only within the conditions and time periods represented in their validation data. It follows that extending validity beyond these bounds demands continuous testing against newly acquired independent data.

3. Calibration and Validation Challenges in Rock Engineering Practice

Three examples, discussed in the following Sections, illustrate the validation challenges rock engineering faces: i) the search for an accurate and precise determination of the Hoek-Brown parameter m_i , ii) the assumption that the rate at which m_b/m_i changes with decreasing rock mass quality is lithology-independent, and iii) the concept of the Representative Elementary Volume (REV).

3.1. Empirical Parameters and the Search for Universal Validation

In 1980, Hoek and Brown [7] provided the following formula as the basis for an empirical criterion to characterize rock mass strength:

$$\sigma_1 = \sigma_3 + \sqrt{m\sigma_c\sigma_3 + s\sigma_c^2} \quad (1)$$

where:

- σ_1 is the maximum principal stress at failure
- σ_3 is the minor principal stress applied to the specimen
- σ_c is the uniaxial compressive strength of the intact rock material in the specimen.

m and s are constants that depend on the properties of the rock mass and the degree to which it has been disturbed or fractured before being subjected to σ_1 and σ_3 .

As the Hoek-Brown criterion evolved (see Equation 2, by [13]), the original parameter m was differentiated into m_i (subscript i for intact rock) and m_b (subscript b for broken or jointed rock mass). In Equation 2, the parameter σ_c was changed to σ_{ci} to represent the uniaxial compressive strength derived from fitting the Hoek-Brown failure envelope to a set of laboratory tests. Equation 2 also includes an exponent a to differentiate between intact rock (a equal to 0.5) and disturbed rock mass through the use of GSI (Geological Strength Index, [14]).

$$\sigma_1 = \sigma_3 + \sigma_{ci} \left(m_b \frac{\sigma_3}{\sigma_{ci}} + s \right)^a \quad (2)$$

$$m_b = m_i \text{Exp} \left(\frac{GSI - 100}{28 - 14D} \right) \quad (3)$$

$$s = \text{Exp} \left(\frac{GSI - 100}{9 - 3D} \right) \quad (4)$$

$$a = \frac{1}{2} + \frac{1}{6} \left(e^{-\frac{GSI}{15}} - e^{-\frac{20}{3}} \right) \quad (5)$$

Practitioners generally agree that laboratory testing provides the most defensible basis for m_i determination. However, these tests are not routinely conducted for most projects. In the literature, there is debate over whether it is reasonable to rely on database correlations as estimates when site-specific data are unavailable [13, 15]. Considerable effort has been devoted to developing precise m_i determination methods [15, 16]. While we agree that testing should be the preferred method to determine m_i , for intact rock samples, the argument against using published values remains mainly focused on intact rock data and massive rock masses [16], as well as highly disturbed rock masses [15]. Extending these conclusions to blocky and very blocky rock masses ($30 < GSI < 70$) represents a significant extrapolation, given the absence of direct field-scale strength measurements and the criterion's sensitivity to GSI variations within this range.

Using data for Middleton Mine [17] and an undisclosed mine location [18], Figure 1 (a, b) shows how the parameter m_i affects the shape of the resulting Hoek–Brown failure envelope, for different GSI (± 5) and s_{ci} (assumed equal to the average UCS). For the two cases under consideration, the influence of m_i variations is most significant at low confining stresses, ranging from 0 to 2 MPa. While this analysis examines the compressive stress region, comparable effects would be anticipated in the tensile regime.

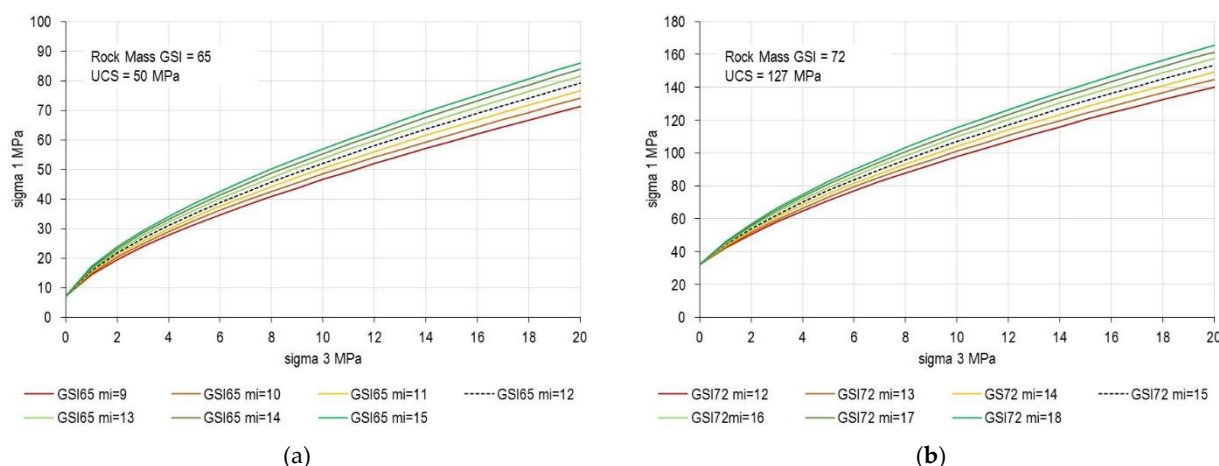


Figure 1. Hoek-Brown curves for two different rock masses as a function of the parameter m_i : (a) UCS = 50 MPa, GSI = 65 and $m_i = 12 \pm 3$, (b) UCS = 127 MPa, GSI = 72, $m_i = 15 \pm 3$.

At the same time, Figure 2 (a, b) shows that s_1 predictions differ by less than 11% between the extreme cases of m_i ($m_i = 12 \pm 3$, and $m_i = 15 \pm 3$, respectively) and assuming that s_{ci} will also vary by $\pm 20\%$. These results suggest that for blocky and very blocky rock masses, the influence of m_i variations may be less pronounced than literature focused on intact rock or massive rock masses would imply. The emphasis on precise m_i determination may therefore stem from the engineering profession's drive to assign numerical precision to geological parameters, rather than from demonstrated sensitivity of design outcomes to m_i values.

These results are a good example of the mathematical optimism that pervades rock engineering practice [19], as if the assumption that adding analytical complexity to empirical relationships will overcome their foundational limitations as site-specific and context-dependent parameters.

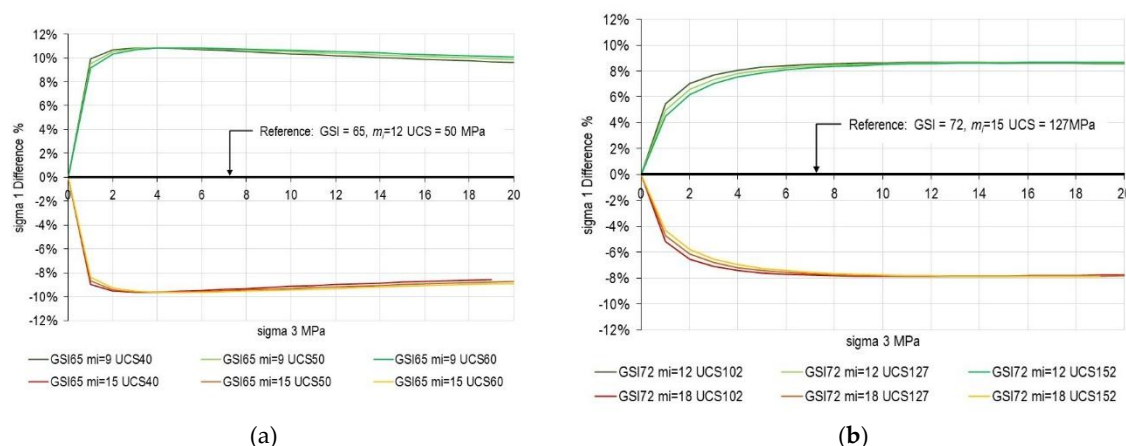


Figure 2. (a) σ_1 percentage difference using UCS = 50 MPa, GSI = 65, and $m_i = 12$ as reference, cases for UCS = 50 ± 10 MPa, GSI = 65, $m_i = 12 \pm 3$, and (b) σ_1 percentage difference using UCS = 127 MPa, GSI = 72, and $m_i = 15$ as reference, cases for UCS = 127 ± 25 MPa, GSI = 72, $m_i = 15 \pm 3$.

What remains less explored is the extent to which accurate m_i estimation affects practical design decisions. Figure 3 demonstrates that pursuing m_i precision offers little practical value. For the Middleton mine rock mass, the range of failure envelopes produced by varying m_i from 9 to 15 ($m_i = 12 \pm 3$) at GSI = 65 can be equivalently reproduced by holding m_i constant at 12 and varying GSI by just ± 5 . Furthermore, the full spectrum of s_i corresponding to the scenarios shown in Figure 2(a) can be reproduced by varying GSI by ± 5 and s_{ci} by 20% while still holding m_i at 12. This range of variability is not unusual in rock engineering practice and is representative of the rock mass at Middleton mine [17, 20]. Figure 3 shows that GSI uncertainty can exert a comparable or greater influence on calculated strength parameters than m_i . This is particularly significant, given that GSI is not a quantitative measurement but a qualitative assessment, which is necessarily subject to observer interpretation and geological judgment.

In this context, the proliferation of GSI quantification methods in the literature [21] may reflect the profession's difficulty in accepting that some parameters resist quantification, regardless of analytical sophistication. Miyoshi et al. [22] illustrated this uncertainty using a DFN model for Middleton mine, showing that GSI estimates varied significantly based on assumptions about the range of in-situ block size distribution and the orientation of the mapped rock face.

Perhaps we should acknowledge that, for blocky to very blocky rock masses ($30 < \text{GSI} < 70$), the precise determination of m_i is subordinate to the degree of natural fracturing, as further discussed in the following Section.

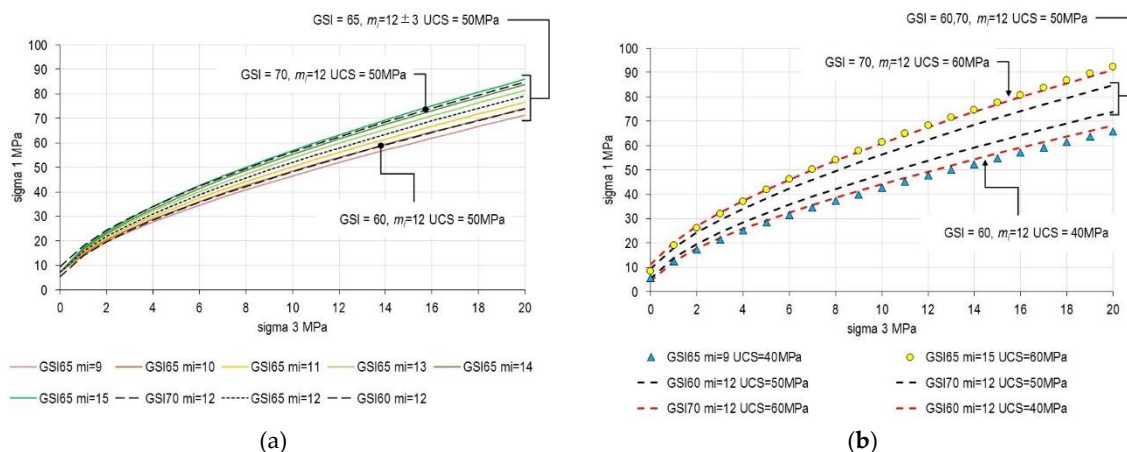


Figure 3. (a) Rock mass conditions corresponding to $\text{UCS} = 50 \text{ MPa}$, $\text{GSI} = 65$, and $m_i = 12 \pm 3$ and $\text{UCS} = 50 \text{ MPa}$, $\text{GSI} = 65 \pm 5$ and $m_i = 12$ (black dashed lines), (b) Rock mass conditions corresponding to $\text{UCS} = 50 \text{ MPa}$, $\text{GSI} = 65 \pm 5$ and $m_i = 12$, and $\text{UCS} = 50 \pm 10 \text{ MPa}$, $\text{GSI} = 65 \pm 5$, and $m_i = 12$ (lines) compared to the extreme range of conditions expected at Middleton mine (triangle and circle symbols).

3.2. From Limited Data to Established Practice

What constitutes accuracy for empirically-derived parameters and for qualitative assessments? Furthermore, with reference to the definitions presented earlier in Table 1, how can parameters be meaningfully calibrated to case studies when their underlying formulations have never been validated through field-scale testing of jointed rock masses and when quantification methods for qualitative indices yield potentially inconsistent results? Addressing these questions requires acknowledging that rock engineering practice frequently relies on “established standards” despite their validation gaps. This raises questions about whether validation requirements are being applied consistently across established and new (emerging) methodologies.

The definitions of m_b and s (Equations 3 and 4) originated from triaxial testing of Panguna andesite (Hoek and Brown, 1980). Despite being derived from a restricted dataset, Equations (3) and [4] are now applied universally across rock engineering practice. However, none of the Panguna andesite specimens represented actual large-scale jointed rock masses, but rather consisted of

reasonably large-scale reconstituted samples (571 mm diameter) or small-scale specimens (152 mm diameter).

The relationship between the m_b/m_i ratio and rock mass quality was inferred from correlations with estimated RMR_{76} [23] and Q values [24]. It was later extrapolated to all rock types beyond the original dataset, with GSI replacing RMR_{76} . Hoek and Brown [7] explicitly acknowledged the scarcity of reliable field-scale testing data needed to validate this extrapolation. Recently, [12] demonstrated that a fundamental issue arises from the assumed universality of Equations (3) and (4), as the same exponential relationship applies to all rock types, meaning different rock masses may experience proportionally identical strength reductions relative to their input m_i and GSI values (Figure 4).

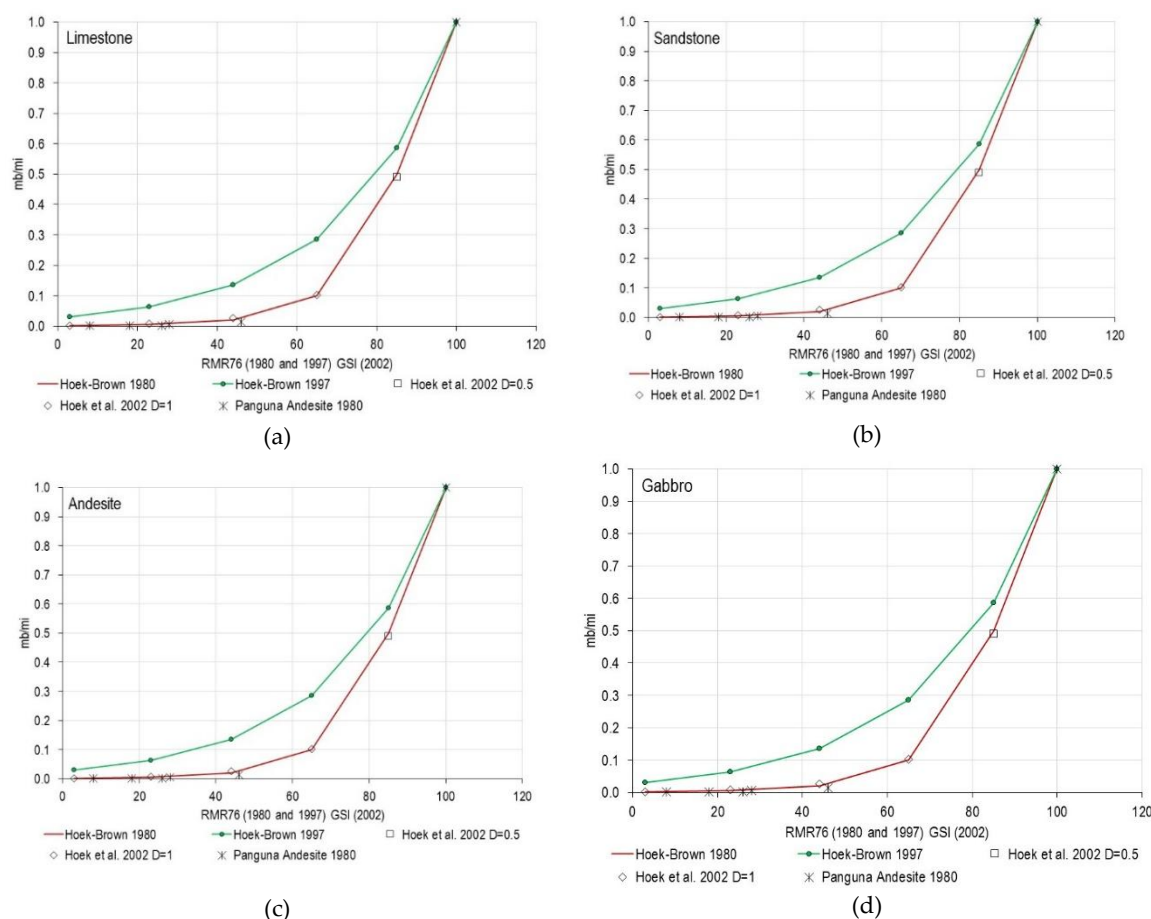


Figure 4. m_b/m_i relationships for different rock types (based on data published originally in Table 12 in Hoek and Brown, 1980). Modified from Elmo and Zoorabadi (2025).

Figure 4 illustrates how the universal empirical relationships (3) and (4) give rise to the “false similarity problem” identified by [25], where fundamentally different rock masses may yield identical strength parameters. For example, Figure 5 shows that while m_b values of 3.86 (GSI = 62, m_i = 15) and 3.79 (GSI = 44, m_i = 30) differ by less than 2%, they represent dramatically different geological conditions. Although the m_b values for the four scenarios shown in Figure 5 are nearly identical, the failure envelopes begin to diverge only once GSI exceeds 80. While Figure 5 might suggest an inverse proportional relationship between m_i and GSI, this apparent correlation is merely an artifact of Equation (3), which mathematically couples these parameters.

Conversely, a practitioner presented with the failure envelopes in Figure 5, and given only UCS and m_i values, would not be able to distinguish whether the underlying rock mass is massive, blocky, or very blocky. This aligns with conclusions by [8], who argue that reducing geological complexity to empirical parameters can remove information that cannot always be reconstructed. The professional implications are profound. Technical reviewers may not be able to work backward from

reported parameters to verify the geological understanding that informed them, limiting the effectiveness of independent review processes.

One might object that Figure 5 presents purely conceptual scenarios and that practitioners should rely on site-specific mapped GSI values and measured m_i parameters. However, this objection exposes a fundamental challenge. Suppose we insist that m_i and GSI must be determined based on specific site conditions because rock masses are too variable for generalized relationships. How can we simultaneously accept that Equations (3) and (4), which transform those site-specific inputs into strength parameters, are universally applicable across all geological conditions and field sites? Clearly, the assumption that for blocky to very blocky rock masses, the parameter m_i , remains independent of natural fracturing deserves scrutiny.

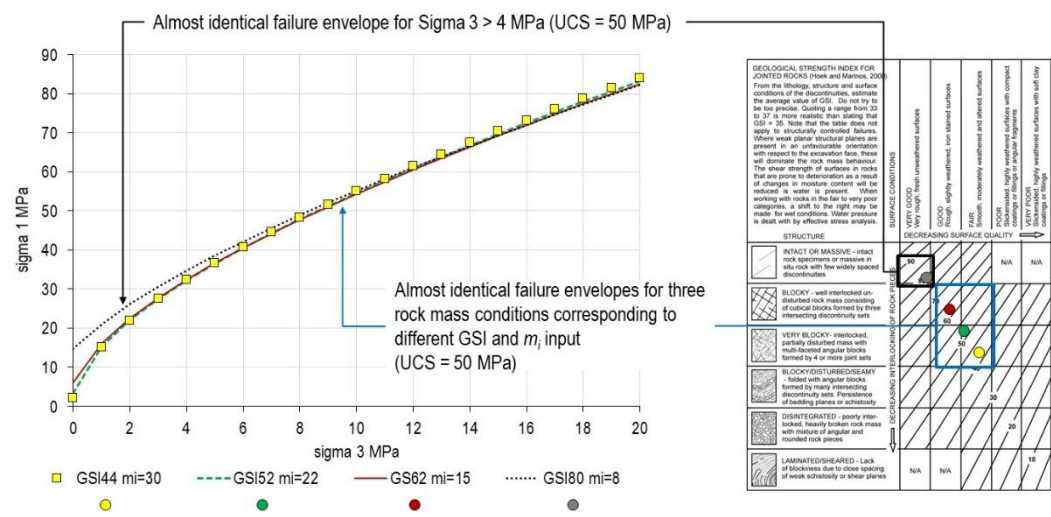


Figure 5. Example of rock mass conditions yielding almost identical Hoek-Brown curves despite representing dramatically different geological conditions.

If m_i relates to fracture mechanics and crack initiation processes [16], one would expect it to vary with fracture connectivity, which governs stress localization through block interlocking mechanisms. Using data from [17], Figure 6 shows that when fitting Hoek-Brown curves to simulated biaxial test data (synthetic rock mass, SRM, models), the derived m_i parameter varies with fracture intensity. While reasonable curve-fitting can be achieved by constraining m_i to the intact rock value ($m_i = 12$) and adjusting only GSI, this approach yields systematically different GSI estimates compared to allowing both parameters to vary (GSI range of [38, 79] vs. [50, 75], respectively, with reference to Figure 6).

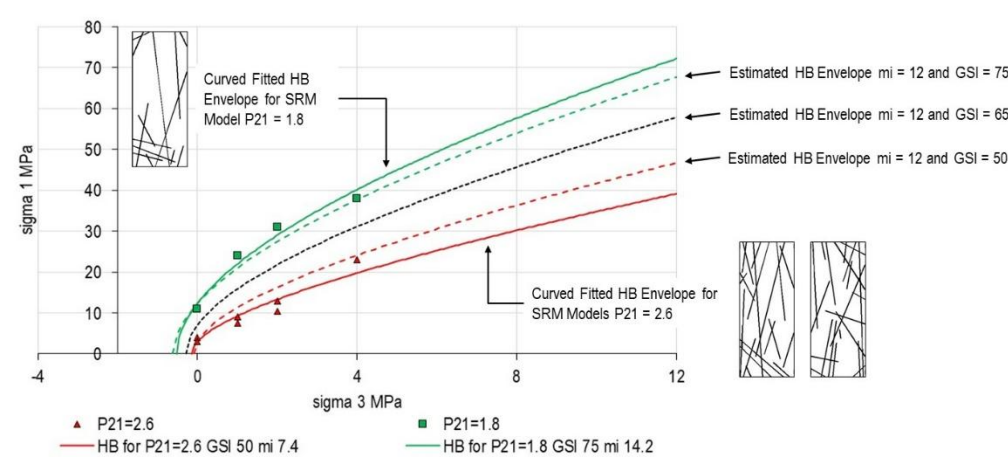


Figure 6. Hoek-Brown curves for SRM models of Middleton mine (2.8 m x 7 m dimensions). Comparison between curves for which both m_i and GSI must be fitted simultaneously to reproduce the simulated behaviour, and curves with equivalent GSI but for which m_i is considered as a material constant independent of structural context ($m_i = 12$ as per intact rock assumption).

This creates a methodological impasse rooted in the false similarity problem discussed above: either i) we treat m_i as a material constant independent of structural context, requiring only GSI calibration to match SRM results, or ii) we acknowledge that both m_i and GSI must be fitted simultaneously to reproduce the simulated behaviour. The former approach (i) derives from the original 1980-1995 conceptual framework, which was developed without large-scale validation, and assumes that m_i remains invariant regardless of fracture network characteristics. The latter (ii) represents a more rigorous interpretation of curve-fitting to simulated results and aligns with the original Hoek-Brown [7] formulation (Equation 1), where the material parameter m should be empirically derived from tests that include structural components.

This analysis exposes two fundamental questions: which parameterization better represents the underlying physics? If m_i varies mechanistically with fracture intensity in SRM models that explicitly simulate brittle fracture processes, should we force it to remain constant to preserve a continuum-based interpretation not supported by these mechanistic simulations?

To address these questions, Figures 7 to 9 present the results of a series of SRM models conducted for three different lithologies (designated as A, B, and C, with reference to the geological units in [18]). Details about the specific nature of the finite-discrete element method (FDEM) approach used in the simulations are provided in [17]. Relevant material properties are listed Table 2 and Table 3. Note that the m_i parameter in Table 2 was determined from triaxial tests of intact rock samples.

Table 2. Material Properties for SRM Models A, B, and C.

Property	Rock Mass A	Rock Mass B	Rock Mass C
Density (ton/m ³)	2.70	2.66	2.61
Uniaxial Compressive strength, UCS (MPa)	67.20	69.9	96.50
Indirect Tension, σ_t (MPa)	2.38	3.07	3.84
Hoek & Brown m_i (laboratory data)	17.36	16.13	20.77
Young's Modulus, E (GPa)	20.06	29.48	37.12
Poisson ratio	0.21	0.21	0.21
Cohesion (MPa)*	9.48	10.16	12.46
Friction angle*	57	57	60
Fracture Energy G_f (J/m ²)	5.97	6.75	8.39

* Calculated in RSData, envelope range for 200 m depth

Table 3. Properties for pre-existing and induced fractures in SRM models A, B and C.

Pre-existing fractures (DFN traces)	Rock Mass A	Rock Mass B	Rock Mass C
Cohesion (MPa)	0.5	0.5	0.5
Friction coefficient (tangent)	0.83	0.83	0.83
Normal Stiffness (GPa/m)	100	50	50
Shear Stiffness (GPa/m)	10	5	5
New fractures properties	Rock Mass A	Rock Mass B	Rock Mass C
Cohesion (MPa)	0.0	0.0	0.0
Friction coefficient (tangent)	0.6	0.6	0.6
Normal Stiffness (GPa/m)	35	25	50

The Hoek-Brown curves for the SRM models in Figures 7 to 9 combine biaxial results from three DFN realizations at the 10th, 50th, and 75th percentile fracture intensities. While the results broadly validate the Hoek-Brown framework, the relationship is not uniform. Rock Mass B, for instance, exhibits a minimal strength difference between the 50th and 75th percentile fracture intensities, consistent with observations by [26] and [27] that fracture intensity (whether areal fracture intensity, P_{21} , or volumetric fracture intensity, P_{32}) is insufficient to characterize rock mass strength across different structural configurations. Rock mass strength and interlocking are fundamentally intertwined. Interlocking describes how block geometry, relative to stress orientation, controls stress transfer pathways and governs strength mobilization. Therefore, fracture intensity alone cannot capture the impact of interlocking because it ignores the mechanical interaction between blocks.

This highlights a fundamental gap between field-based GSI and SRM-derived GSI estimates. Field-based GSI values transform the qualitative assessment of rock mass structure and surface conditions into equivalent continuum properties via Equation (2); therefore, they do not account for interlocking directly. Conversely, the SRM-derived GSI values mechanistically reflect how block geometry relative to stress orientation mobilises strength through mechanical interlocking.

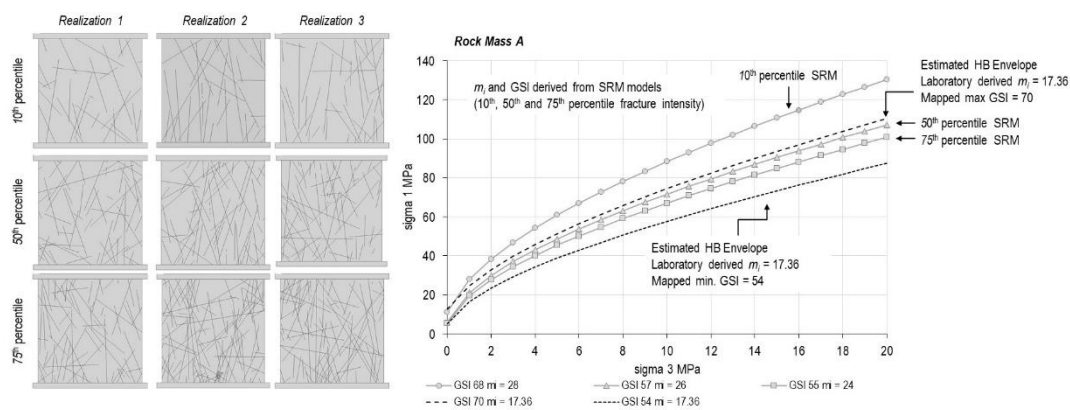


Figure 7. Hoek-Brown curves derived based on a series of biaxial tests for Rock Mass A (Elmo et al., 2020) for different fracture intensities and multiple DFN realizations. The Hoek-Brown curves corresponding to the intact rock m_i value and the mapped GSI are also included.

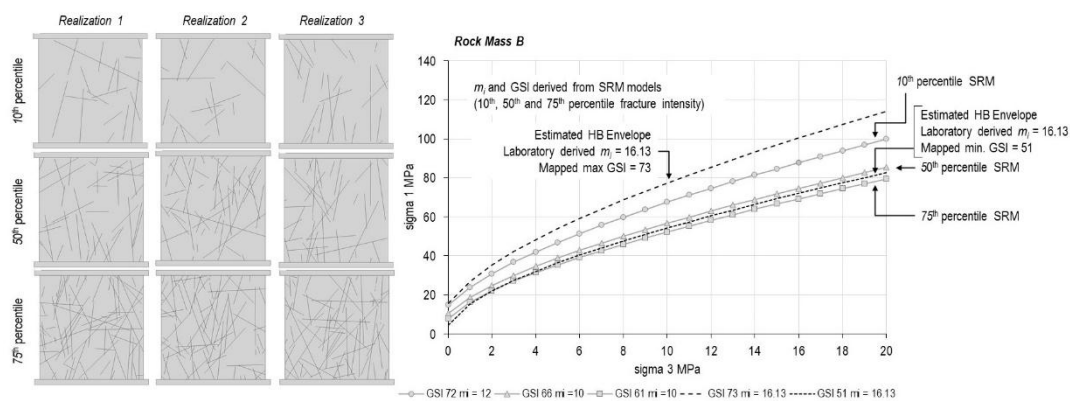


Figure 8. Hoek-Brown curves derived based on a series of biaxial tests for Rock Mass B (Elmo et al., 2020) for different fracture intensities and multiple DFN realizations. The Hoek-Brown curves corresponding to the intact rock m_i value and the mapped GSI are also included.

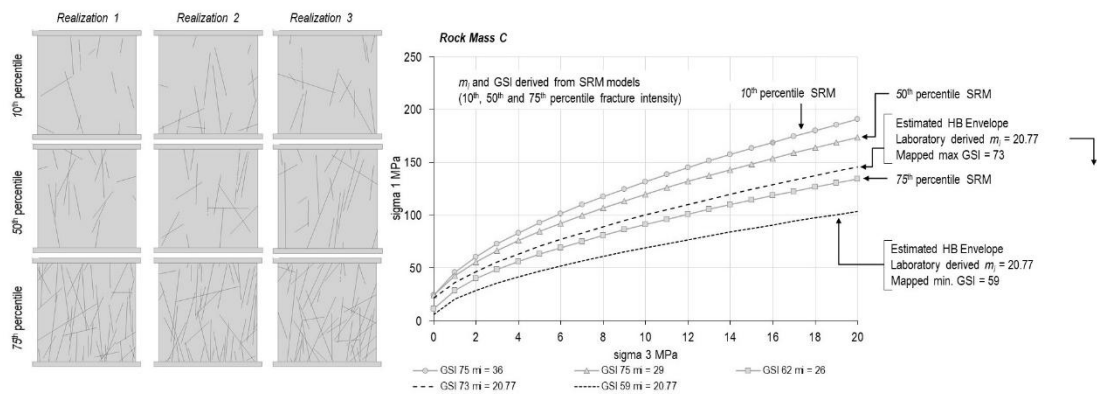


Figure 9. Hoek-Brown curves derived based on a series of biaxial tests for Rock Mass C (Elmo et al., 2020) for different fracture intensities and multiple DFN realizations. The Hoek-Brown curves corresponding to the intact rock m_i value and the mapped GSI are also included.

Figures 10 compare field-based GSI with SRM-derived GSI estimates using the two fitting approaches introduced earlier: i) GSI fitted with m_i constrained to laboratory values, and ii) GSI and m_i fitted simultaneously. The former systematically performs worse than the latter when compared to field-based GSI values. The results thus suggest that treating m_i as independent of rock mass structure (blocky to very blocky conditions) requires reconsideration.

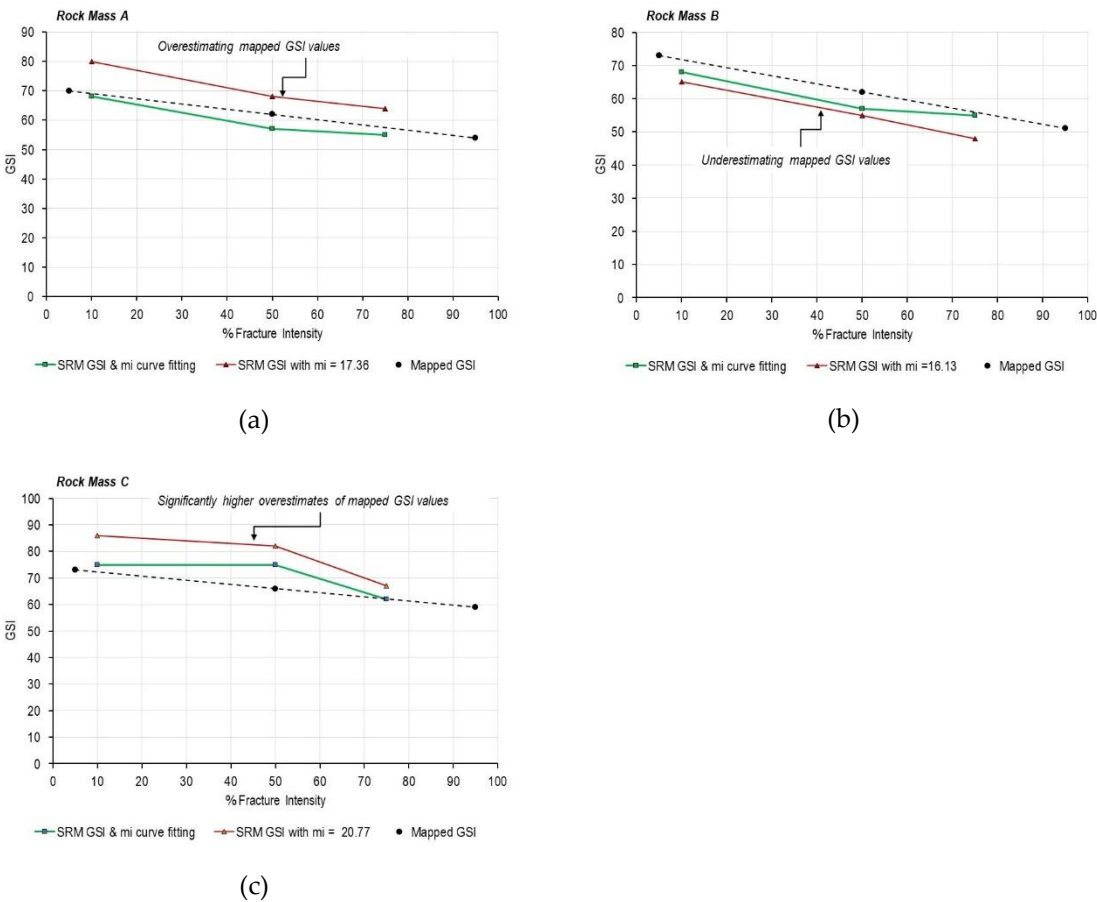


Figure 10. Comparison between field-based GSI (dashed line) with SRM-derived GSI estimates using the two fitting approaches introduced in the text: i) GSI fitted with m_i constrained to laboratory values (red line), and ii) GSI and m_i fitted simultaneously (green line).

To close this Section, it is important to note that the realism of SRM models depends on the accuracy of the underlying discrete fracture network (DFN). In terms of geometrical representation of the natural fracture network, SRM results inherit all uncertainties and variabilities present in the underlying DFN models. Moreover, because DFN generation is stochastic, SRM analysis becomes inherently probabilistic. This is a characteristic with important implications for engineering design that traditionally expects deterministic outcomes.

Critically, field-based GSI values are not exempt from uncertainty, as they are subject to qualitative judgment and rely on the observer's interpretation, geological experience, and the inherent difficulty of transforming qualitative assessments into quantitative values. The question is not whether field-based GSI values are "accurate" and SRM-derived values are "inaccurate" or vice versa. Instead, both approaches carry inherent uncertainties. Field-based GSI values, often derived from core logging rather than mapping of rock exposures, reflect a subjective assessment of rock mass quality. In contrast, SRM-derived GSI values reflect uncertainties in DFN representation and mechanical modelling. The objective of either approach should be establishing a reliable range of GSI conditions that brackets plausible behaviour, rather than pursuing illusory precision.

The probabilistic reality of the SRM models shown in Figures 7 to 9 illustrates a pattern identified earlier in Section 3.1. Geological complexity resists reduction to deterministic parameters, whether through empirical quantification schemes (GSI) or computational sophistication (DFN-based SRM models). The profession's repeated attempts to eliminate uncertainty through methodological elaboration, whether by quantifying GSI, refining m_i determination methods, and increasing DFN model complexity, reflect the difficulty in accepting that some geological characteristics remain fundamentally indeterminate regardless of analytical investment. The relationship between uncertainty and rock engineering knowledge is explored further in Sections 4 and 7.

3.3. The Representative Elementary Volume: Conceptual Limitations and Validation Challenges

The Representative Elementary Volume (REV) concept, widely used in rock engineering practice, is purported to establish the dimensions above which rock mass properties become scale-invariant. However, this interpretation fundamentally misconstrues what REV represents and conflates distinct concepts (structural homogenization, size effects, and mechanical behaviour) into a single, convenient abstraction.

The conventional understanding suggests that REV defines a critical volume beyond which rock mass strength stabilizes, implying that larger volumes exhibit consistent properties while smaller volumes show scale-dependent variation. This interpretation has become so ingrained that REV and "size effects" are now treated as synonymous in much of the literature. Yet this equivalence rests on a conceptual interpretation of what is actually being observed in numerical models.

Classical illustrations [13] and modelling of REV (e.g., [28, 29, 30, 31, 32]) suggest that as the problem scale increases, rock mass behaviour becomes less sensitive to the presence of adversely oriented individual discontinuities. This observation pertains to the statistical averaging of structural variability, rather than fundamental changes in material strength. As initially conceived by Bear [33] for porous media, REV represents the volume at which statistical homogeneity emerges, not a threshold beyond which physical properties fundamentally change.

In his discussion of scale effects, Pinto da Cunha [34] referred to geometrically homothetical samples of the same rock or rock mass, subjected to similar loading conditions. The use of the appropriate terminology merits clarification. Scale effects refer to mechanical characteristics that change when a system is proportionally altered across different hierarchical levels, maintaining geometric similarity (homothetic samples). Size effects, conversely, refer to how absolute dimensions influence behaviour independent of proportionality. In rock engineering discourse, "size effects" remains the prevalent terminology for three practical reasons [27]: i) the dimensional difference between laboratory specimens and in-situ rock masses is substantial rather than incremental; ii) the literature typically presents strength as a function of absolute dimensions rather than dimensionless scaling ratios; and iii) rock engineering problems are not necessarily scalable. For example, pillar

dimensions are dictated by orebody geometry, design constraints, and economic considerations rather than simple geometric proportionality.

Central to understanding scale effects is the concept of homothety, that is, a mathematical transformation that proportionally scales all geometric features of an object about a fixed center point by a constant ratio k [35]. However, the term "geometrical homothety" introduces a crucial distinction that is systematically overlooked in rock engineering applications. True geometric homothety requires that the scaling function apply not only to external dimensions but to all structural features at all scales, including mineral grains, microcracks, macroscopic discontinuities, and the statistical distributions governing their occurrence. A genuinely homothetical rock mass model of twice the linear dimension would contain discontinuities of twice the length, spaced at twice the interval, with twice the aperture, embedded in a matrix with grain sizes doubled and microdefect spacing doubled accordingly. Real engineering problems do not meet these conditions.

SRM and DFN studies claiming to demonstrate size effects apply the homothety function exclusively to the external boundaries. This represents partial homothety. The model boundaries are scaled up, but the structural representation continues to obey its own spatial distribution function. This distinction is not purely linguistic, as significant effects emerge when critically oriented structures are included or excluded stochastically as the model volume changes, introducing variability unrelated to scale. Similarly, attempts to define the REV of a rock mass using only geometric descriptors (e.g., volumetric fracture intensity P_{32} , and volumetric fracture count, P_{30}) decouple structure from mechanism, ignoring how fracture networks translate into mechanical behaviour. Geometric scale-invariance does not guarantee mechanical scale-invariance because geometric characterization methods (e.g., the vertical axis in the GSI table, P_{32} , P_{30} , etc.) ignore loading conditions, which fundamentally influence how structure translates into mechanical behaviour. These methods establish isotropic conditions a priori.

While SRM studies are often used to demonstrate apparent size effects (as modelled volumes decrease, calculated rock mass strength increases), this interpretation may confuse mechanisms with dimensions [27]. As model dimensions change, the proportion of the rock mass influenced by discontinuities relative to intact rock varies, fundamentally altering whether behaviour is dominated by joint mechanics or intact rock fracture. These changes represent alterations in structural configuration rather than pure size effects. The observed strength variations with model dimension reflect the mechanistic consequences of different structural realisations, rather than an intrinsic scale and size dependency of rock mass strength. The problem is compounded by geometric effects. Anisotropic strength varies not only with size but also with specimen shape, meaning different size effect relationships can be derived in principle depending on the model geometry [36]. More critically, using fully homothetical SRM models, Bewick and Elmo [27] demonstrated behaviour contrary to conventional size effect theory, in which rock mass strength either remains scale-invariant (models without DFNs) or actually decreases with decreasing size (models with homothetical DFNs).

Applied to rock masses, this means that REV addresses the conditions under which continuum approximations become statistically reasonable, rather than the scale or size at which rock mass strength increases or decreases [37]. The concept of rock mass complexity, introduced by [37], fundamentally concerns the limits of applicability of homogenized failure criteria, such as the Hoek-Brown criterion, rather than actual physical changes in rock mass strength with volume. This discussion reveals an additional paradox of the validation of rock engineering problems raised in Section 2. We cannot validate scale effect relationships in rock masses because there are no truly homothetical rock mass samples at the field scale. Laboratory studies can approach full homothety only within narrow size ranges and only for intact rock. Field observations necessarily sample different structural configurations rather than scaled versions of identical structures.

4. The Scientific Limitations of Empirical Methods

When studying rock mass behaviour, we assume a straightforward process: i) collect data, ii) make observations, and iii) synthesize interpretations. This assumes theory-neutral observation,

which is epistemologically naive. Without prior frameworks, how do we determine which observations are significant, which data to collect, and at what scale? Observation is never theory-neutral; we always bring conceptual frameworks that guide both our search strategies and interpretive processes.

When mathematical and validation frameworks prove insufficient, as established in Section 3, expert judgment is often invoked in rock engineering practice to fill the gap. This judgment represents a form of implicit knowledge that cannot be easily formalized or transmitted through equations. However, expert judgment introduces subjective elements that contradict Popper's criterion for scientific knowledge [8]. This problem becomes acute when experts disagree about geological conditions or describe the same rock mass using different material parameters (e.g., Figures 5 and 6). Such disagreements reveal that our rock engineering knowledge may be more accurately described as informed opinion shaped by individual experience. Expert judgment becomes a necessary but epistemologically problematic component of rock engineering practice.

In this context, we can illustrate rock engineering knowledge as a cone that reduces uncertainty by capturing it (Figure 11). Rock mass knowledge thus is akin to studying the interior of a dark room through scattered pinholes (boreholes) and small surface openings (outcrops), inferring a three-dimensional structure from limited one-dimensional views. Data collection efforts aim to cut larger openings, yet can never fully illuminate the interior. The metaphor emphasizes the difference between uncertainty, which diminishes with adequate data access (once sufficient data exist, additional observations do not materially change understanding or conclusions), and radical uncertainty [38], which persists because critical features may remain undetected, not due to inadequate sampling but because they represent unknown unknowns. A single such observation can disproportionately alter aggregate knowledge regardless of sample size. Therefore, fundamental limitations will always remain in making rational decisions. This requires shifting to a different epistemological goal, one that does not seek accuracy but seeks to manage uncertainty. Yet uncertainty itself is not what prevents us from labelling rock engineering knowledge as "scientific." For rock engineering knowledge to qualify as truly "scientific", it would have to follow the principle by [10] that [quote] "*past observations may lead to the discovery of a theory, but the theory must predict the future*". Paraphrasing the same author, even when our models fit existing data, they must incorporate a mathematical framework that makes them predictive over time for rock engineering knowledge to be scientifically valid.

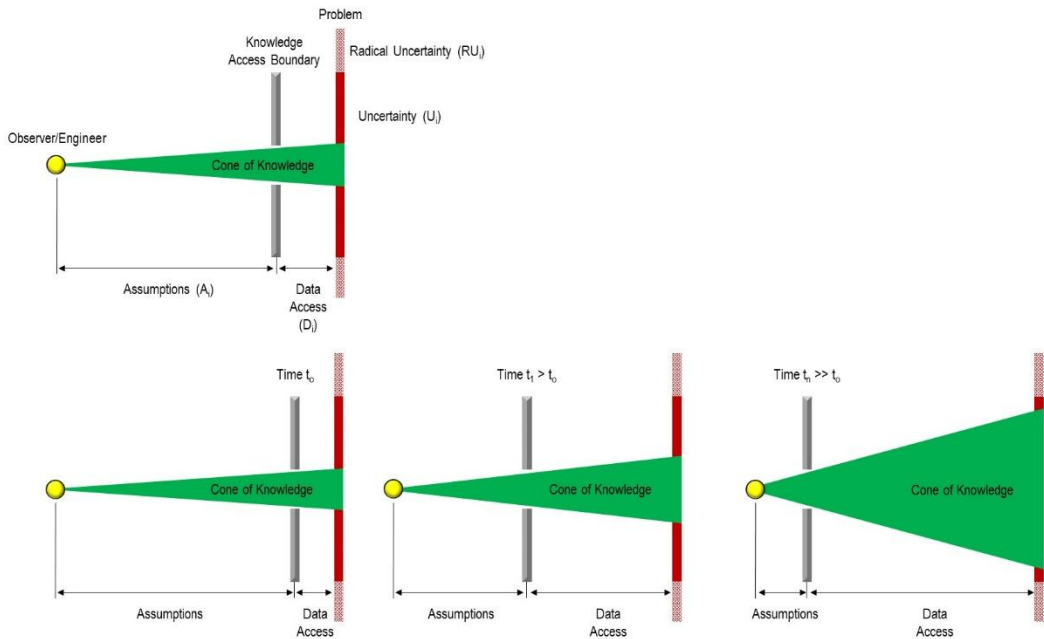


Figure 11. Rock engineering knowledge is imagined as a cone that reduces uncertainty by capturing it. Note the distinction between uncertainty and radical uncertainty.

Additionally, since our empirical criteria can only be verified through observation, the practical constraints on conducting field-scale experiments necessarily limit our capacity to fully validate theoretical frameworks such as the Hoek-Brown criterion (Sections 3.1 and 3.2). Following Einstein's [39] epistemological principles, theories require empirical validation rather than acceptance on purely a priori grounds. As discussed in Sections 2 and 3, many methods commonly used in rock engineering practice derive much of their legitimacy from their historical role in the evolution of rock engineering. These systems have earned their position through decades of application and refinement. However, from an epistemological perspective, their continued validity depends on ongoing empirical evaluation rather than historical precedent alone. This suggests that their use should be accompanied by a critical assessment of their applicability to contemporary conditions and emerging challenges. Otherwise, the transition from "credible guides" to "industry standards" would merely confirm our profession's resistance to critical examination [8].

The same principle highlights a fundamental limitation in rock engineering practice, as our models and empirical methods often fail to achieve genuine predictive capability. Rather than addressing this fundamental limitation, we continue to add more data points to our databases and update our empirical criteria. But this process provides no certainty about future predictions, regardless of the accuracy we claim for the empirical parameters used in our design calculations. This is what Taleb [40] referred to as the turkey problem, which illustrates the danger of inferring future outcomes from past patterns. It is as if somehow the models and empirical methods used in rock engineering design were pointing at the future through a backward-facing process (Figure 12).

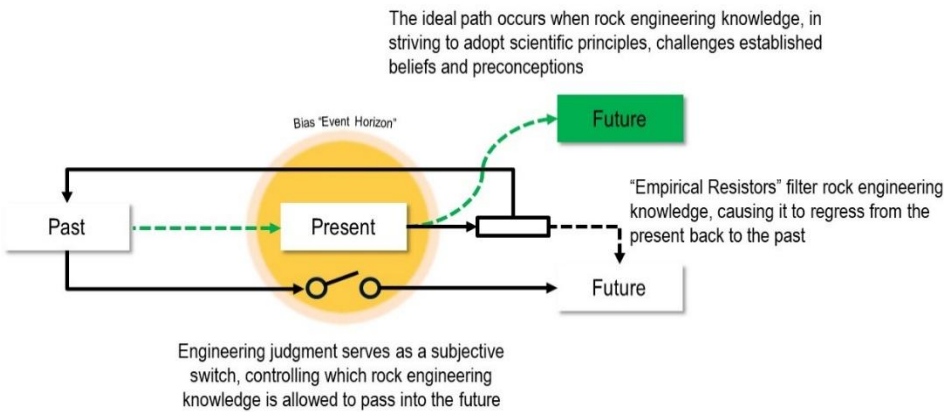


Figure 12. Impact of cognitive biases and engineering judgment on how rock engineering knowledge moves into the future by first regressing into the past.

This process reveals the limitations of the classical problem of induction. The assumption that future rock mass behaviour will resemble past behaviour cannot be proven, but it is accepted as a pragmatic necessity. This places rock engineering knowledge on uncertain epistemological foundations, regardless of the mathematical sophistication we claim to adopt. Nonetheless, this limitation may not be decisive, since, given the inherent uncertainty we initially identified (Sections 2 and 3), rock engineering predictions should never be framed in terms of certainty. Under absolute certainty, risk disappears entirely [38]. Either we possess complete knowledge, enabling failure-proof design, or we know that failure is inevitable, regardless of intervention, prompting alternative design approaches.

5. Rock Engineering and The Challenge of Operational Definitions

Bridgman [41] challenged how scientists conceptualize measurement and meaning. Bridgman argued that scientific concepts should be defined entirely through the operations used to measure them, which he named "operational definitions." For Bridgman, the concept of length, for example, means nothing more than the set of operations by which length is measured. This operationalism emerged from his recognition that classical physics had relied on concepts that could not be operationally defined, resulting in conceptual confusion. For rock engineering models to qualify as scientific theories, they would require operational definitions that link model symbols to measurable physical events, as well as the capacity for quantitative predictions of future events [10].

Applied to rock engineering, Bridgman's operationalism reveals profound challenges. Consider the concept of rock mass strength. What operations define this concept? We cannot directly measure rock mass strength in situ without destroying the very structure we seek to analyze. Instead, we infer rock mass strength through indirect measurements and observations, including combinations of laboratory tests of intact rock samples, back-analysis of failures, and empirical correlations with classification indices (Section 3). However, each of these operations measures and observes something different. Laboratory tests measure the strength of intact rock under controlled conditions that deviate from the field stress states. Back-analysis assumes our numerical models can correctly simulate actual failure mechanisms. Classification systems attempt to quantify qualitative assessments through a series of rating schemes. And yet, even when combined, none of these operations offers a direct measure of rock mass strength.

The disconnect between measurement and reality deepens when one considers that, in rock engineering, we routinely use terms such as rock quality, joint roughness, and weathering degree as if they were measurable physical properties. However, the operations defining these concepts often involve subjective visual assessments, arbitrary numerical scales, or proxy measurements that may not capture the underlying physical phenomena we intend to characterize. Similarly, the Rock Quality Designation (RQD, [42]), operationally defines rock quality as the percentage of core pieces greater than 10 cm in length. This operational definition says nothing about actual mechanical quality. It measures only one aspect of fracture spacing as observed in core samples. Yet, RQD has become accepted as an indicator of rock mass quality and incorporated into strength calculations as if it directly measures an actual mechanical property, notwithstanding its reliance on an arbitrary 10 cm spacing threshold [43, 11].

The operational definitions framework requires that conceptual models be verified against the physical operations that define our measurements. This creates what we might call the verification paradox since conceptual failure models can only be verified through observed failures, which is precisely what engineering design seeks to prevent. For instance, our conceptual models of slope failure incorporate assumptions about failure surfaces, strength parameters, and triggering mechanisms. To verify these models operationally, we would need to observe actual slope failures under controlled conditions. However, such verification would require either:

- Deliberately inducing failures in prototype slopes (ethically and practically unacceptable)
- Waiting for natural failures to occur (temporally impractical for design purposes)
- Relying on historical failures (which introduces temporal and contextual uncertainties)

This paradox means that our most critical design concepts cannot be operationally verified. Nonetheless, when rock engineering models successfully reproduce past failures, we often interpret this as validation. This exemplifies a *post hoc ergo propter hoc* fallacy, which assumes that because models match past failures, they can predict those failures. Bridgman's operationalism rejects such reasoning because it conflates correlation with causation and retrofitting with prediction. Accurate operational verification would require that models, developed independently of the phenomena they claim to explain, successfully predict future observations [38]. In rock engineering, this standard is rarely met. This pattern exemplifies the broader calibration and validation challenge discussed in Section 2, where rock engineering exhibits the weakest discrimination between these fundamentally different processes. Instead, we adjust model parameters until they reproduce known outcomes, then assume these calibrated models possess predictive validity.

For rock engineering models to qualify as scientific theories in Bridgman's sense, they would require operational definitions that link model symbols to specific, reproducible measurement procedures, as well as a demonstrated capacity for quantitative future predictions that are independent of the data used in model development. By this criterion, most empirical methods in rock engineering practice fail to achieve a genuine scientific status (e.g., the imposition of a 10cm threshold for RQD, discussed by [11]). They remain useful engineering tools, although they fall short of the standards set by scientific theories.

When one considers a given rock engineering problem, our duty is to reflect on the phenomena and create a model of it using a mathematical framework. Note that the mathematical framework in our case is not necessarily created upon direct measurements of physical properties. As such, the description of the phenomena may lead to different outcomes. GSI is a perfect example of this problem (Section 3.1). More than 25 tables and formulations have been proposed to quantify GSI [21]. And yet, all those tables and formulations can yield different results. As discussed in Section 2, the notion of validating any of those GSI quantification methods proves problematic. Since validation criteria vary across methodological frameworks, the same method can simultaneously be validated and invalidated depending on the evaluative lens applied. A better approach would be to ask ourselves why we need multiple methods of quantifying GSI when none eliminates its inherently qualitative nature. On the contrary, these quantification methods are typically validated by comparing their numerical outputs against qualitative GSI assessments. This represents circular reasoning, as we create quantitative methods to avoid subjectivity, then validate them against subjective assessments.

Experience remains the criterion upon which we assess the applicability of our design methods, including GSI. However, creative principles and innovation reside in the ability to challenge what the experience dictates us to use. Similarly, when deploying Hoek-Brown and GSI, we are implicitly accepting that rock engineering design can be based on deterministic methods and under the assumption of a homogeneous equivalent continuum medium. But in practice, as discussed in Section 3.2, whether we consider intact rock samples or rock masses at the field scale, there is always epistemic uncertainty and aleatoric variability [44], combined with tacit subjectivity regarding the values of any parameter.

5.1. Pragmatic Operationalism and Risk Assessment

The authors do not argue for dismantling six decades of rock engineering practice or declaring established methods invalid, despite their inherent subjectivity and limitations. Rather than abandoning current practice, we suggest adopting what could be called "pragmatic operationalism", explicitly acknowledging that our measurements are proxies rather than direct observations of the phenomena we seek to understand. This approach would emphasize understanding the limitations and assumptions embedded in our operational definitions rather than treating them as windows onto physical reality.

This approach does not necessarily undermine risk characterization, though it may reframe what risk quantification means from a rock engineering perspective. Rather than treating factors of safety or probabilities of failure as precise predictions, pragmatic operationalism recognises them as estimates bounded by the limitations of our proxy measurements and empirical correlations. In this context, an evaluation of confidence refers to an assessment of the degree of uncertainty, rather than certainty. For example, when using a factor of safety approach, we tend to use a larger margin, not because we are certain of our input, but because we acknowledge the uncertainty of our estimated input or its variability. Accordingly, a factor of safety of 1.3 based on well-validated methods within their calibration range carries different epistemic weight than a factor of safety of 1.5 derived from extrapolated empirical relationships.

Figure 13 shows that radical uncertainty persists in geological systems and cannot be truly eliminated. It represents an ontological feature of complex geological systems rather than a limitation of our current knowledge state [8]). More importantly, from a risk perspective, radical uncertainty

manifests in the form of unknown unknowns, and as such, it cannot be managed using statistical methods. Similarly, Unforeseen conditions often become the default explanation when reality diverges from prediction, but this attribution carries significant implications depending on what type of uncertainty these unforeseen conditions represent:

- If they represent uncertainty (whether epistemic or aleatoric), this means that these conditions are unforeseen due to inadequate data collection and characterization. From a legal perspective, this amounts to admitting we failed to recognize that we did not collect sufficient information.
- If they represent radical uncertainty, this means no one could claim they would have acted differently, since the conditions would not have been known to them either.

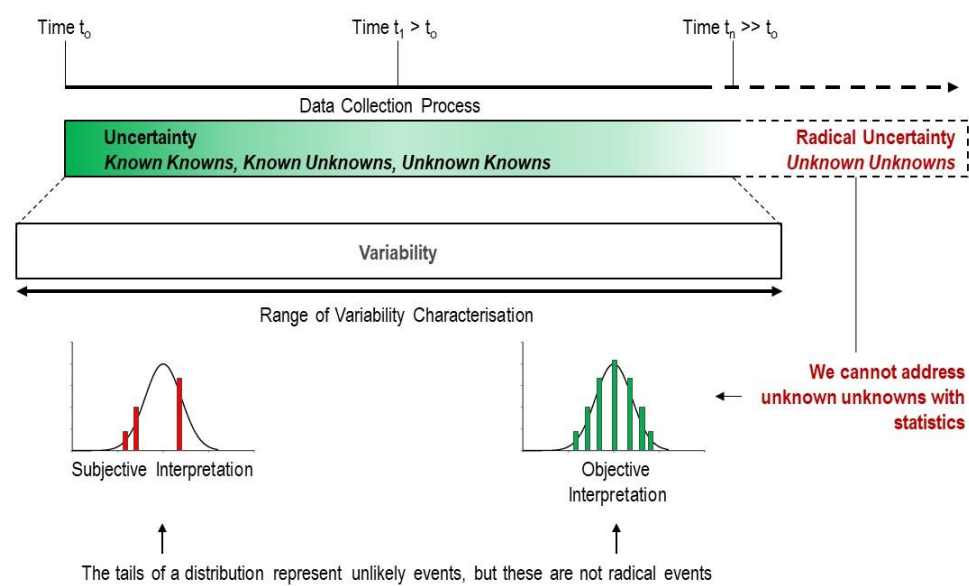


Figure 13. The difference between uncertainty and radical uncertainty and implications for statistical characterization.

6. The Epistemological Limits of Modelling

Throughout this paper, we have examined various numerical and empirical approaches, from SRM models to classification-based strength estimation. The fundamental question is not whether numerical models can help us determine rock mass strength, but instead whether they provide genuinely helpful information within the context of radical uncertainty.

Rock mass strength does not exist as a singular, deterministic value waiting to be discovered through increasingly sophisticated analysis [45]. Instead, it manifests as a range of conditional responses that vary with scale, stress path, and time-dependent processes. This range reflects not only measurement uncertainty but also the inherent variability of natural rock mass systems. A critical challenge underlies all rock mass modelling, as it remains fundamentally impossible to verify a priori whether our results accurately represent a future reality. Unlike laboratory specimens, which can be subjected to strength testing at scales relevant to engineering applications, we cannot directly test rock masses or build prototypes of our excavations. Even when physical models are constructed at considerable expense, like the case of the model used prior to the Vajont Dam disaster of 1963, they may offer no advantage over numerical models when founded on incorrect assumptions.

If we do not know the present, we certainly do not know the future. Models, therefore, serve as tools to expose what we do not know. Given these limitations, models should primarily serve as tools for exploring potential conditions, while acknowledging the boundaries of our knowledge. Their highest value lies not in providing definitive answers, but in systematically confronting us with what we do not and cannot know with certainty. The role of models should be to challenge our

preconceptions and force explicit consideration of alternative scenarios, rather than reinforcing existing beliefs through seemingly sophisticated analysis.

However, models may fail to redirect our thinking if we unconsciously guide them toward conclusions we have already reached through other means. This represents a subtle but pervasive form of confirmation bias in rock engineering practice [8]. When model inputs, boundary conditions, and constitutive relationships are selected based on implicit expectations about outcomes, the resulting analysis becomes a sophisticated form of circular reasoning rather than genuine investigation.

When expanding the argument to artificial intelligence (AI), the challenge of using complex models or advanced mathematics leads us to believe that we can fit a large number of parameters to our data. However, while the fitting may look reasonable, the model may lack a mechanistic representation and thus may lack predictive ability. As our models become increasingly complex, experimental studies become even more imperative. We drown in modelling results, but we are starved of actual field-scale testing.

6.1. The Mechanism Selection Paradox

Consider the fundamental contradiction in how we approach stability analysis in practice. Engineers face two seemingly similar but epistemologically distinct scenarios:

- Scenario A: Modellers are asked to conduct stability analysis because the governing failure mechanism is unknown.
- Scenario B: Modellers are asked to conduct stability analysis to help identify the governing failure mechanism.

The two scenarios reveal that linguistic framing can mask fundamental philosophical equivalence. The choice of numerical method should ideally be based on the likely failure mechanisms. Yet, we often cannot know these mechanisms until after the analysis is complete, or until post-excavation monitoring provides empirical evidence of actual behaviour. This creates a methodological circular argument where the tool selection depends on knowledge that the tool itself is meant to generate.

Scenario A treats the model as a substitution for missing knowledge using numerical analysis to provide answers despite our ignorance of the governing failure mechanisms. These answers may contradict each other; furthermore, their realism depends entirely on human interpretation, and there is no guarantee that engineers and practitioners will always agree on a given interpretation. Scenario B treats the model as an investigative tool for discovery, systematically exploring failure mechanisms to gain a deeper understanding of the system. However, Scenario B suffers from the same fundamental limitation as Scenario A. We cannot determine a priori which failure mechanism is correct. Scenario B adopts a more optimistic linguistic tone, yet produces the same epistemological outcome as Scenario A. Whether we refer to it as "substituting for knowledge" (Scenario A) or "investigative discovery" (Scenario B), we still end up with the same linguistic hedging, the same subjectivity for the modelling inputs, and the same fundamental inability to definitively identify the governing failure mechanisms.

The simplification from real rock mass to equivalent continuum models further complicates the problem. Figure 14 illustrates the progression from a real rock mass scenario to an equivalent continuum model (Figure 14). As models become simplified, from realistic, undulating fractures through increasingly simplified DFN representations to fully isotropic continuum models, mechanistic accuracy decreases. More importantly, as we simplify toward homogeneous continuum models, we move further from reality. How can we validate a fully isotropic continuum model against a highly anisotropic, discontinuous rock mass? Matching observed outcomes through calibration proves only that parameters can be adjusted to fit data, not that the simplified model captures the actual failure mechanism. We could argue that the more we abstract away from mechanistic reality, the less meaningful our validation processes become.

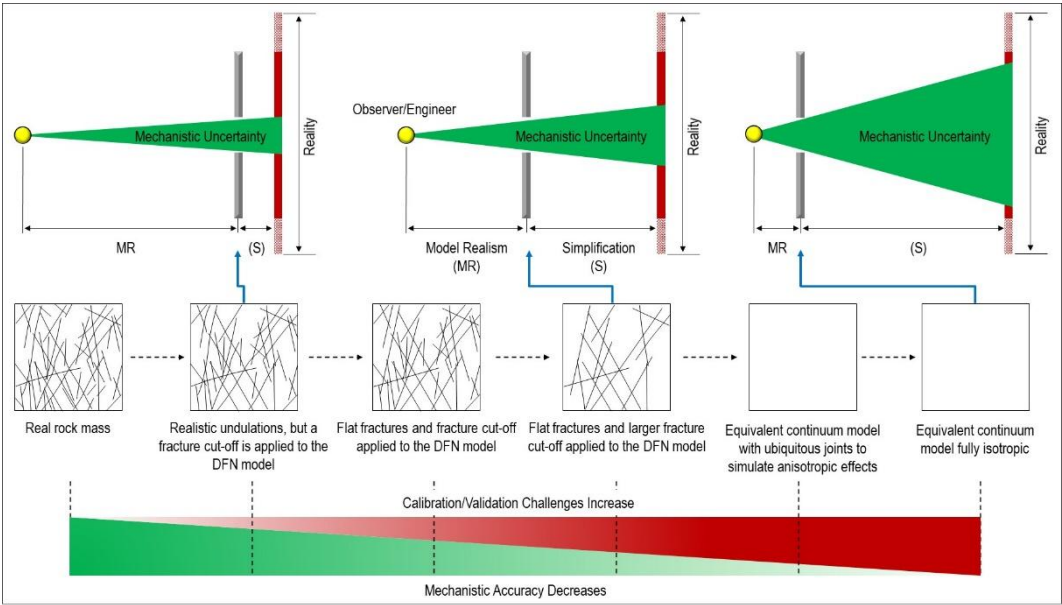


Figure 14. The validation paradox in rock mass modelling. As models progress from mechanistic realism (MR) to simplification (S), calibration/validation challenges increase while mechanistic accuracy decreases.

7. Uncertainty and Professional Responsibility in Rock Engineering Practice

Sections 2 through 6 have revealed significant challenges in rock engineering knowledge. We now address important practical implications. The challenge for practitioners becomes managing professional responsibility and risk while acknowledging that our most widely used empirical and numerical methods often produce results that cannot be validated against field-scale reality. At the same time, rock engineering practice demands apparent certainty, driven by pressure related to professional liability and safety considerations. Consequently, empirical methods such as GSI, RMR, and Q, which are intended to characterize the variability of rock mass conditions, are instead used as quantitative and supposedly precise measurement tools to derive rock mass properties.

The Large Open Pit (LOP) guidelines [46] illustrate how rock engineering practice prescribes probability thresholds that represent epistemic uncertainty rather than actual failure frequency. Kay and King [38] would refer to these probability thresholds as subjective probability. For instance, a 50% probability of failure threshold for bench design does not predict that half of the benches will fail; instead, it acknowledges that design can proceed despite knowledge limitations, with the understanding that operational management and monitoring will complement analytical predictions. They represent a form of pragmatic operationalism already embedded in our design practice. We proceed with structured professional judgment despite radical uncertainty, relying on operational experience, observational methods, and adaptive management to bridge the gap between our analytical predictions and complex geological reality. In the authors’ opinion, making this pragmatism explicit, rather than hiding it behind the search for seemingly precise calculations (see Sections 2 and 3), could serve the profession better.

7.1. The Challenge of Dual Uncertainty

When rolling a die, we face uncertainty about the outcome but possess certainty about the constraints, as the result will necessarily fall between 1 and 6. This represents bounded uncertainty. We do not know the specific outcome, but we are aware of the possibilities and their associated probabilities. Rock engineering offers no such epistemological comfort as it confronts what might be termed dual uncertainty:

- First, we remain uncertain about our inputs. What are the actual rock mass properties at depth? How do joint properties vary spatially? What is the true in-situ stress state? These input

uncertainties reflect not only measurement limitations but also fundamental constraints on observing three-dimensional geological structures through one-dimensional sampling.

- Second, even if we could magically eliminate all input uncertainty and know exact geological conditions, we would still face output uncertainty: What will actually happen when we excavate? Which failure mechanism will dominate? How will the rock mass respond to changing stress conditions over time? Will progressive failure occur, and if so, at what rate? This output uncertainty exists because geological systems exhibit emergent behaviour, scale-dependent mechanisms, and time-dependent processes that cannot be fully predicted even with perfect knowledge of the initial conditions.

Dual uncertainty distinguishes geological systems from most other engineering domains. A structural engineer designing a steel frame faces input uncertainty about material properties and loading conditions, but given those inputs, the output behaviour is well-constrained by established theory. The steel's yield strength may vary within a range, but its fundamental mechanical behaviour is known and reproducible. In contrast, rock mass behaviour involves applying empirically derived approximations of complex, scale-dependent phenomena to uncertain parameters. The constitutive laws we use remain empirical. Furthermore, the interaction between input and output uncertainty creates non-linear epistemological effects. Small uncertainties in input parameters can produce disproportionately large uncertainties in predicted outcomes, not just through sensitivity analysis but through qualitative changes in failure mode. A rock mass that appears stable under one set of assumptions within our uncertainty bounds might fail catastrophically under a slightly different but equally plausible set of assumptions.

7.2. Professional practice vs. Uncertainty and Radical Uncertainty

This distinction between uncertainty and radical uncertainty raised in Section 5 has profound implications for engineering practice. The standard paradigm of "reduce uncertainty through better measurement" has fundamental limits in rock engineering practice. No amount of additional boreholes can eliminate the fact that we are sampling a three-dimensional, heterogeneous geological body through one-dimensional windows. No improvement in laboratory testing can resolve scale-dependent mechanisms where small specimens do not represent large rock masses. And no refinement of numerical models can overcome our uncertainty about both the inputs we provide and the mechanisms by which those inputs are translated into outputs. Our models attempt to predict rock mass behaviour as if unknown knowledge follows the identical distributions we have established for known knowledge, as illustrated earlier in Figure 13. But if we cannot fully comprehend the present state of a rock mass, our predictions about its future behaviour rest on even more unstable foundations.

7.3. Professional Practice vs. Linguistic

Linguistic imprecision compounds the challenges discussed in previous Sections. Claims of "accurate rock mass strength" or "accurate determination" of empirical parameters pervade rock engineering practice and the literature. Yet, claims of accurate rock mass strength require closeness to an actual value that cannot be measured a priori; at best, we can model rock mass behaviour and derive an apparent strength, but this approach encounters the same validation challenges described in Section 3. Where the "true value" of geological parameters cannot be determined independently, claims of accuracy represent a category error. We cannot be accurate about quantities that lack objective physical meaning or cannot be measured in the field (e.g., rock bridges [26]).

The term "reliable" adds another layer of complexity. Reliability refers to a system or process that consistently yields predictable and repeatable results. Critically, our design methods can be reliable even if they are not entirely accurate. For example, empirical strength formulae are often reliable because engineers using the same formula and inputs consistently obtain the same result. Yet, as discussed in Section 3, empirical relationships carry uncertainty due to being extrapolated beyond the limited calibration data and specific field conditions. Figure 15 maps this problem in a

reliability-uncertainty space. At the beginning of any project, rock engineering practice occupies the upper-right quadrant (R_1U_2), characterized by high reliability but high uncertainty. In principle, given a set of initial estimates, we can reliably determine the rock mass strength and, with it, calculate a factor of safety; however, the validity of our assumptions means that this reliable calculation may be consistently incorrect. The ideal state (quadrant R_2U_2 , lower right) requires both procedural consistency and epistemic confidence in the form of adequate data access and validated assumptions. As shown earlier in Figure 11, our cone of rock engineering knowledge will not capture the entire set of uncertainty, and radical uncertainty persists as a second layer superimposed on the reliability-uncertainty space.

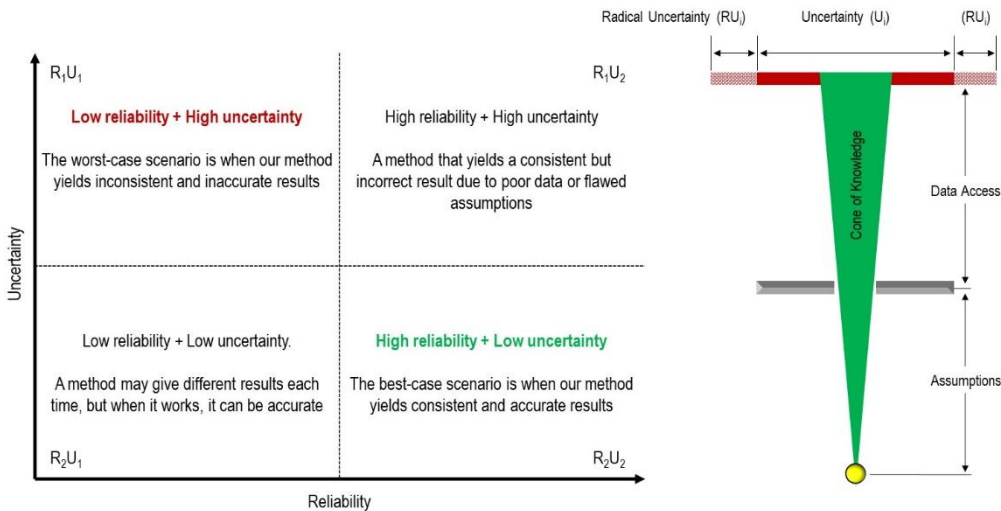


Figure 15. Reliability-uncertainty space. The cone of knowledge illustrates how uncertainty diminishes with access to data. However, radical uncertainty persists beyond data-driven reduction.

7.4. The Algorithmic Amplification of Uncertainty

Professional and research practices are increasingly seeking to add an even more complex layer through AI-driven characterization methods, potentially compounding rather than resolving these foundational epistemological challenges. Tolstoy's observation [47] that [quote] “all that we know is that we know nothing” captures a fundamental problem in rock engineering practice. While theoretically we accumulate vast databases of data, rework our empirical correlations, and develop increasingly sophisticated numerical models, each apparent advance in knowledge can reveal new depths of geological complexity, and it does not protect our design from radical uncertainty.

The present value of this quote is revealed in earlier work on a priori vs. a posteriori knowledge with the limits of computational predictions by [8]. Our numerical models represent the manifestation of what we believe we know. Yet, their inability to predict unknown unknowns (radical uncertainty) suggests that our confidence in computational tools may itself be a form of cognitive bias. As we increasingly delegate engineering judgment to algorithmic systems, Tolstoy's paradox becomes more relevant than ever. When dealing with incompletely validated models and methods, encoding fundamentally limited human knowledge into systems that appear cognitively superior to humans creates an epistemological illusion, in which algorithmic complexity appears to transcend the cognitive limitations in the underlying models and methods, as previously discussed in Sections 2 and 3.

8. Conclusions

Rock engineering practice faces significant challenges, including the conflation of calibration with validation (Section 2) and the use of empirical relationships that lack validation against field-scale experiments (Section 3). Additionally, it encounters radical uncertainty and the limitations of

operational definitions (Sections 4-6). These challenges create profound professional implications (Section 7).

A quote attributed to Valery Legasov [48], the scientist who oversaw the Chernobyl nuclear disaster, states that [quote] *"To be a scientist is to be naïve. We are so focused on our search for the truth, we fail to consider how few actually want us to find it."* Applied to rock engineering problems, this quote reflects both practical difficulties (e.g., the challenges of large-scale testing of jointed rock masses) and an understandable professional reluctance to undertake studies whose outcomes might require fundamental reassessment of established practice. The absence of systematic validation for commonly adopted design methods and practices may suggest an inadvertent approach that favours maintaining established methods over subjecting them to the same validation protocols demanded of new approaches. Validation protocols should apply uniformly. If new methodologies require validation across diverse conditions, established methods should also be subjected to ongoing empirical scrutiny and not accepted on an a priori basis, particularly when they are applied beyond their original context.

A significant contribution of this paper is to demonstrate how SRM models (Section 3) validate the Hoek-Brown framework, as expressed by Equation (1), while simultaneously revealing important distinctions between theoretical formulation and practical applications:

- Equation (1) does not require GSI, and the parameters m and s emerge from fitting the results of a series of triaxial tests (or biaxial SRM models, as in our paper).
- Equation (2) requires GSI, from which m_b and s are calculated as functions of the assumed GSI value.

While we could argue that Equation (2) represents a form of pragmatic operationalism (Section 5.1), it is important to recognize that the assumed GSI and calculated m_b and s in Equation (2) can differ significantly from the emergent GSI, m and s parameters from fitting Equation (1) directly to a series of SRM models.

The resulting false similarity problem exemplifies the uncertainty themes developed throughout this paper:

- The search for "accurate" and "precise" parameter determination (Section 2) proves misguided when different parameter combinations produce equivalent outcomes. For instance, the combined Hoek-Brown and GSI approach should be considered in the context of homogenizing and simplifying a jointed rock mass into an equivalent continuum medium.
- GSI quantification methods (Section 3) cannot resolve this ambiguity, as no quantification scheme can eliminate the non-uniqueness inherent in the GSI framework itself.
- Additional testing and characterization may reduce uncertainty about which parameterization to use, but cannot eliminate radical uncertainty.

A second major contribution of this paper is articulating principles for responsible practice in the face of uncertainty. Rather than pursuing precision in individual parameters, responsible rock engineering practice should acknowledge the false similarity problem that may affect our empirical methods, focusing more on mechanisms rather than parameter precision. The proliferation of AI-driven tools also requires particular caution. Algorithms trained on subjective assessments and unvalidated correlations cannot transcend the epistemological limitations of their training data. As discussed in this paper, computational sophistication risks creating false confidence in parameters that remain fundamentally non-unique and empirically unvalidated.

Rock engineering necessarily operates under a standard of "balance of probabilities" (Section 2). Given the nature of radical uncertainty, professional practice should evaluate methods not by whether they eliminate uncertainty, but by whether they represent reasonable approaches to managing it. The path forward requires what might be termed "epistemological integrity", i.e., aligning our claims with our actual knowledge, our professional communications with our methodological limitations, and our validation demands with consistent principles. This does not mean abandoning current practice but rather practicing with explicit acknowledgment of its empirical foundations.

Author Contributions: Conceptualization, D.E. and S.A.; methodology, D.E. and S.A.; formal analysis, D.E. and S.A.; investigation, X.X.; resources, D.E.; data curation, D.E. and S.A.; writing—original draft preparation, D.E.; writing—review and editing, D.E. and S.A.; visualization, D.E.; supervision, D.E.; All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research received no external funding.

Data Availability Statement: No new data were created in this study. Modelling results presented in Section 3 are based on papers by the first author. These are included in the references.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
GSI	Geological Strength Index
RMR	Rock Mass Rating
DFN	Discrete Fracture Network
SRM	Synthetic Rock Mass
REV	Representative Elementary Volume
AI	Artificial Intelligence
ML	Machine Learning

References

1. Stevens, S.S. On the theory of scales of measurement. *Science* 1946, 103, 677–680.
2. Morgenroth, J.; Unterlaß, P.J.; Sapronova, A.; Khan, U.T.; Perras, M.A.; Erharter, G.H.; Marcher, T. Practical recommendations for machine learning in underground rock engineering—On algorithm development, data balancing, and input variable selection. *Geomech. Tunn.* 2022, 15, 650–657.
3. Editorial on Papers Using Numerical Methods, Artificial Intelligence and Machine Learning. *Rock Mech Rock Eng.* 2023, 56, 1619 <https://doi.org/10.1007/s00603-023-03278-3>
4. Yang, B. Examining the reliability of integrating machine learning with rock mass characterization and classification data, Doctoral thesis, 2024, University of British Columbia, Vancouver.
5. Elmo, D. 'The risk of confusing model calibration and model validation with model acceptance. in PM Dight (ed.), SSIM 2023: Third International Slope Stability in Mining Conference, Australian Centre for Geomechanics, Perth, pp. 333–342.
6. Flexner, A. The usefulness of useless knowledge., *Essay, Harpers*, Issue 179, June/November 1939.
7. Hoek, E.; Brown, E.T. *Underground excavations in rock*. London: Instn Min. Metall., 1980.
8. Elmo, D.; Stead, D. The Role of Behavioural Factors and Cognitive Biases in Rock Engineering. *Rock Mech. Rock Eng.* 2021, 54, 2109–2128
9. Popper, K. *The Logic of Scientific Discovery*. Routledge, 1959, pp. 545.
10. Dougherty, E.R. *The evolution of scientific knowledge: from certainty to uncertainty*. Bellingham, Washington: SPIE Press, 2016.
11. Yang, B.; Mitelman, A.; Elmo, D.; Stead, D. Why the future of rock mass classification systems requires revisiting its empirical past. *Q. J. Rock Eng. Hydrogeol.* 2021, 55, qjeh2021-039.
12. Elmo, D.; Zoorabadi, M. Examining the case for accuracy and precision when determining the Hoek-Brown parameters. Submitted to *Rock Mech. Rock Eng.*, 2025. Under Review.
13. Hoek, E.; Brown, E. Practical estimates of rock mass strength. *Int. J. Rock Mech. Min.* 1997, 34, 1165–1186
14. Hoek, E. Strength of rock and rock masses. *ISRM News J.* 1994, 2, 4–16.
15. Marinos, V.; Carter, T. Integrating GSI and mi for Reliable Rockmass Strength Estimation: Integrating GSI and mi. *Rock Mech. Rock Eng.*, 2025, 58, 11217–11260.

16. Cai, M. 2010. Practical Estimates of Tensile Strength and Hoek–Brown Strength Parameter m_i of Brittle Rocks. *Rock Mech. Rock Eng.*, 2010, 43. 167–184. 10.1007/s00603-009-0053-1.
17. Elmo, D. Evaluation of a Hybrid FEM/DEM Approach for Determination of Rock Mass Strength Using a Combination of Discontinuity Mapping and Fracture Mechanics Modelling, with Particular Emphasis on Modelling of Jointed Pillars. Doctoral Dissertation, University of Exeter, Exeter, UK, 2006
18. Elmo, D.; Yang, B.; Stead, D.; Rogers, S. A discrete fracture network approach to rock mass classification. In *Challenges and Innovations in Geomechanics*, Proceedings of the 16th International Conference of IACMAG- Volume 1, Turin, Italy, 5–8 May 2021; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 854–861.
19. Barton, N. *Reflections on Unrealistic Continuum Modelling*; NB&A: Oslo, Norway, 2025; 30p.
20. Pine, R.; Harrison, J.P. Rock mass properties for engineering design. *Quarterly Journal of Engineering Geology and Hydrogeology*. 2003, 36. 5–16. 10.1144/1470-923601-031.
21. Yang, B.; Elmo, D. Why engineers should not attempt to quantify GSI. *Geosciences* 2022, 12, 417.
22. Miyoshi, T.; Elmo, D.; Rogers, S. Influence of data analysis when exploiting DFN model representation in the application of rock mass classification systems. *Journal of Rock Mechanics and Geotechnical Engineering*, 2018, 10, Issue 6, 1046–1062,
23. Bieniawski, Z.T. Rock mass classification in rock engineering. In *Exploration for Rock Engineering*; Bieniawski, Z.T., Ed.; Balkema: Cape Town, South Africa, 1976; pp. 97–106.
24. Barton, N.; Lien, R.; Lunde, J. Engineering classification of rock masses for the design of tunnel support. *Rock Mech.* 1974, 6, 189–236.
25. Hadjigeorgiou, J.; Harrison, J.P.. Uncertainty and sources of error in rock engineering. *Harmonising Rock Engineering and the Environment - Proceedings of the 12th ISRM International Congress on Rock Mechanics*. 2012, 2063–2067. 10.1201/b11646-393.
26. Elmo, D. The Bologna Interpretation of Rock Bridges. *Geosciences* 2023, 13, 33.
27. Bewick, R.P.; Elmo, D. Size effect and rock mass strength. *Can. Geotech. J.* 2025, 62, 1–18.
28. Esmaili, K.; Hadjigeorgiou, J.; Grenon, M. Estimating geometrical and mechanical REV, based on synthetic rock mass models at Brunswick Mine. *International Journal of Rock Mechanics and Mining Sciences*, 2010, 47: 915–926, doi:10.1016/j.ijrmms.2010.05.010
29. Elmo, D. 2012. FDEM & DFN modelling and applications to rock engineering problems. Faculty of Engineering, Turin University, Italy (MIR 2012 - XIV Ciclo di Conferenze di Meccanica e Ingegneria delle Rocce— —Nuovi metodi di indagine, monitoraggio e modellazione degli ammassi rocciosi). November 21st - 22nd, 2012.
30. Stavrou, A.; Vazaios, I.; Murphy, W.; Vlachopoulos, N. Refined approaches for estimating the strength of rock blocks. *Geotechnical and Geological Engineering*, 2019, 37: 5409–5439(2019). doi:10.1007/s10706-019-00989-9.
31. Li, Y.; Wang, R.; Chen, J.; Zhang, Z.; Li, K.; Han, K. Scale dependency and anisotropy of mechanical properties of jointed rock masses: insights from a numerical study. *Bulletin of Engineering Geology and the Environment*, 2023, 82: 114. doi:10.1007/s10064-023-03150-2
32. Yiouta-Mitra, P.; Dimitriadis, G.; Nomikos, P. Size effect on triaxial strength of randomly fractured rockmass with discrete fracture network. *Bulletin of Engineering Geology and the Environment*, 2023, 82:8. doi:10.1007/s10064-022-03031-0.
33. Bear, J. *Dynamics of Fluids in Porous Media*; Courier Corporation: North Chelmsford, MA, USA, 2013.
34. Pinto da Cunha, A. *Scale Effects in Rock Masses*. London CRC Press, 1993, pp366.
35. Bewick, R.P.; Elmo, D. Does size effect matter? In *proceedings of the RockEng 2025 Conference*, August 2025, Montreal, Canada.
36. Elmo, D.; Stead, D. An integrated numerical modelling–discrete fracture network approach applied to the characterisation of rock mass strength of naturally fractured pillars. *Rock Mech. Rock Eng.* 2010, 43, 3–19.
37. Erharter, G.; Elmo, D. 2025. Is Complexity the Answer to the Continuum vs. Discontinuum Question in Rock Engineering?. *Rock Mech. Rock Eng.* 1–19. 10.1007/s00603-025-04777-1.
38. Kay J.; King, M. *Radical Uncertainty: Decision-Making Beyond the Numbers*. WW Norton, 2020, pp 384.

39. Einstein, A. Autobiographical notes, in Albert Einstein: Philosopher Scientist. 1959, Schilpp, P. A. (ed.) NY, Harper and Row.
40. Taleb, N. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2010, pp 400
41. Bridgman, P. W. *The Logic of Modern Physics*. 1927, New York: Macmillan.
42. Deere, D.U.; Hendron, A.J.; Patton, F.D.; Cording, E.J. Design of surface and near-surface construction in rock. In Proceedings of the Failure and Breakage of Rock-Eighth Symposium on Rock Mechanics, Minneapolis, MN, USA, 15–17 September 1967; pp. 237–302.
43. Pells, P.J.; Bieniawski, Z.T.; Hencher, S.R.; Pells, S.E. Rock quality designation (RQD): Time to rest in peace. *Can. Geotech. J.* 2017, 54, 825–834.
44. Harrison, J.P. Rock engineering design and the evolution of Eurocode 7. In Proceedings of the EG50 2017 — Engineering Geology 50 Conference, Portsmouth, UK, 5–7 July 2017
45. Ambah E.; Elmo, D.; Znag, Y. Fracture Undulation Modelling in Discontinuum Analysis: Implications for Rock-Mass Strength Assessment. *Geotechnics* 2025, 5(3), 58; <https://doi.org/10.3390/geotechnics5030058>.
46. Read, J.; Stacey, P. 2009. *Guidelines for Open Pit Slope Design*. 2009, CSIRO Publishing, Melbourne.
47. Tolstoy, L. *War and Peace*. Wordsworth Editions, 1993.
48. Legasov. V. Personal audio files. Transcribed by HBO, Chernobyl, 1988.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.