Revisiting probability of failure versus factor of safety – a screening tool with limitations

J J Moreno[1] and S R Kendall[2]

1. Principal Consultant, Tailex Pty Ltd, Perth WA 6018. Email: pepe.moreno@tailex.com.au
2. Senior Consultant, Tailex Pty Ltd, Perth WA 6018. Email: sam.kendall@tailex.com.au

ABSTRACT

The use of simplified methods to establish a relationship between factors of safety and probability of failure (Silva, Lambe & Marr, 2008) has become more popular in tailings dam risk assessments as they offer a probabilistic perspective on stability that can be achieved with relative ease. While valuable as screening tools, their limitations warrant caution, especially when used to assess a TSF's ALARP position. This paper applies a probabilistic analysis to hypothetical cases, aiming to identify whether simplified methods could lead to misleading risk interpretations. A discussion on the limitations of Silva (2008) is presented, and alternative pathways are explored.

INTRODUCTION

As part of evolving industry standards and heightened stakeholder expectations, mining companies are increasingly required to quantify risks related to tailings storage facilities (TSFs), not only to inform engineering decisions but also to support financial reporting through disclosure of potential failure liabilities. A key principle in this context is the ALARP (As Low As Reasonably Practicable) concept, which requires that risks be reduced to a level that achieves balance between safety, practicability and cost.

To demonstrate a TSF's ALARP position and enable clear communication of risk, a TSF's probability of failure (PoF) must be quantified across all relevant failure modes. Slope instability is a commonly observed failure mode (Stark, Moya & Lin, 2022) and often a major contributor to the PoF. To quantify TSF instability risk, practitioners rely on a range of methods, both subjective and objective, each offering strengths and limitations. These limitations must be diligently reviewed when comparing a TSF's ALARP position to tolerable risk thresholds: even risks considered well below societal limits carry potential for misrepresentation.

STATE OF PRACTICE

The current state of practice in assessing TSF stability typically relies on deterministic methods, using either limit equilibrium (LE) or stress-strain numerical modelling techniques, where a calculated safety factor (FoS) is measured against generic minimum criteria.

While industry guidance advocates the use of risk-based approaches, minimum FoS criteria are widely used, including by the authors, reflecting client/regulatory expectations and practical limitations in data availability to inform probabilistic analysis.

Slope stability is influenced by aleatoric uncertainty (arising from natural material variability) and epistemic uncertainty (arising from knowledge gaps or errors in understanding or measurement). Deterministic stability analyses typically use lower bound or characteristic strength values in addition to FoS criteria to manage aleatoric and epistemic uncertainty. This approach does not offer quantifiable indication of instability risk, as a compliant deterministic FoS could be underestimating risk and failing to meet tolerability thresholds—or overestimating risk, resulting in unnecessary conservatism and overdesign.

However, endorsed by GISTM, the industry is increasingly using risk—and PoF—as a means of communicating with stakeholders. Silva (2008) (referred to herein as SLM) offers a practical bridge between deterministic analysis and PoF, allowing an annualised PoF to be inferred from empirical relationships.

Alternatives to SLM include more rigorous objective analyses; however, their complexity, cost and data requirements have impeded their widespread adoption in geotechnical engineering. While these approaches can offer additional precision in certain circumstances, they can also mask uncertainty and dilute accountability.

LIMITATIONS OF SLM

SLM presents a semi-empirical relationship that converts deterministic FoS into annualised PoF estimates. The underlying PoF estimates are informed by expert judgment using details from 75 projects. Notably, the method assumes the input FoS is derived from effective stress analyses and refines PoF estimates by categorising facilities (Class I–IV) according to the quality of investigation, design, construction and monitoring.

The SLM method is a well-founded approach that places value on expert judgment and geotechnical insight. It has useful applications, including screening-level assessments and portfolio risk comparisons. However, its application to quantifying TSF risk to inform TSFs' ALARP positions requires careful scrutiny This view is based on the following notable limitations.

Limited calibration for TSFs: The original development of SLM drew on a broad range of slope types but included limited calibration specifically for TSFs (two case histories), reducing its reliability in this context. SLM does not account for the wide range of TSF configurations—such as integrated waste landforms (IWLs) and filtered tailings stacks—that are often designed with operational objectives beyond stability, including material disposal efficiency or waste haulage/tailings transport logistics.

Loading conditions: The method assumes the input FoS is derived from effective stress analyses. Modern TSF design standards seek to address undrained loading conditions where stability is influenced by contractive materials (such as hydraulically placed tailings or weak foundations).

Foundation conditions: SLM treats foundation conditions indirectly through the input FoS and selection of engineering classes (I–IV) but is unable to explicitly assess or model complex foundation behaviours or heterogeneity—common in foundations over alluvial, colluvial, or highly weathered residual soils.

Conservatism in deterministic design: Standard deterministic design approaches often lead designers to employ lower bound or characteristic strength values to establish the deterministic FoS, which yield a higher PoF from the SLM curves (SLM notes that FoS using mean parameters should be applied). This can result in misguided concern and unnecessary expenditure.

QUANTIFYING POF SENSITIVITY TO SHEAR STRENGTH VARIABILITY

Uncertainty in shear strength is often the largest source of variability in slope stability analyses (Duncan, Wrigth & Brandon, 2014). This paper explores how PoF varies across a realistic range of material shear strength and the coefficient of variation (CoV) values, where the CoV captures aleatoric and epistemic strength uncertainties.

PoF has been estimated for various loading conditions and TSF configurations, as shown in **Figure 1**. Each TSF configuration featured a 30 m embankment height, 1V:3H downstream slopes and a 10 m crest width. The phreatic surface was modelled to simulate a steady-state seepage condition from a supernatant pond that is initiated 100 m upstream of the crest and exits at the downstream toe. Each scenario was assessed considering:

- competent (high strength) and weak (low undrained strength) foundation conditions
- drained strength for all tailings
- undrained strength for all tailings
- drained strength for tailings above, and undrained strength for tailings below the phreatic surface.

A probabilistic modelling approach using GeoStudio software was employed, applying the limit equilibrium method (Morgenstern-Price) with circular failure surfaces. Monte Carlo simulations with 10 000 samples were initially trialled to ensured convergence; however, it was observed that Latin Hypercube Sampling (LHS) with 1000 samples provided comparable accuracy with reduced computational time. Simulations were performed to compute a probability distribution of the FoS for a range of scenarios. The PoF was then estimated from fitted distributions where the FoS falls below unity or P[FoS < 1] (Harr, 1977; Christian, Ladd & Baecher, 1994; Baecher, 1984).

In steady-state conditions without a clear trigger or time-dependent degradation, the probability of failure from probabilistic slope stability analysis reflects uncertainty in the inputs—not the likelihood of failure over time. It is a measure of confidence in the design, not a prediction of failure rate. For this study, a 50-year period was assumed to represent the time during which the TSF remains in steady-state conditions prior to closure and reclamation. Under this assumption, the annual PoF can be approximated by dividing the total PoF by the selected time window. Key to this analysis is assessing the probability of drained and undrained behaviour and estimating their shear strength distributions using values derived from literature. A discussion of these aspects follows.

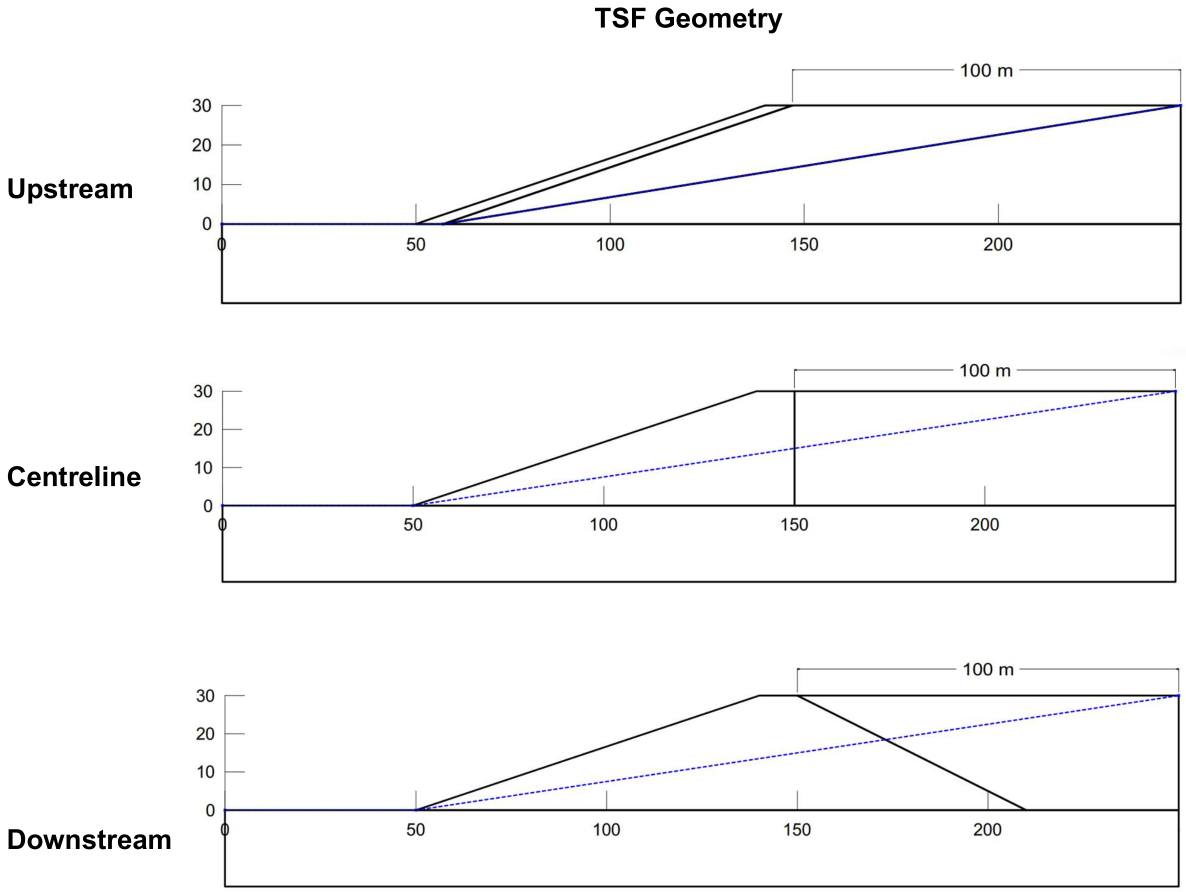

Figure 1: Summary of analysed TSF stability geometries

Drained versus undrained conditions

To capture a realistic range of potential failure mechanisms in TSF slopes, the analysis considers both drained and undrained conditions.

Under drained conditions, failure probability is largely governed by variability in effective strength parameters. Strength uncertainty in undrained scenarios is similarly present, though the mobilisation of undrained strength is often conditional on triggering events, including rapid loading or a consistent rise in phreatic surface.

To account for the conditional nature of mobilising undrained strengths, the LEM PoF was multiplied by the probability of a triggering event, using Barneich *et al*'s (1996) qualitative-to-quantitative mapping. This approach introduces the effect of time dependency, since the probability of triggering would be estimated on a yearly basis. The likelihood of triggering was considered to vary according to TSF status as follows:

- Certain: Actively raised TSFs (eg upstream, rapid rise) approach a trigger probability of 1.0, which is consistent with ANCOLD (2019).
- Unlikely: Intermittent deposition may reduce the probability to $\sim 10^{-2}$.
- Rare: Care and maintenance conditions may reduce probability further, to $\sim 10^{-3}$.

Probabilistic shear strength distributions

The analysis considers three material domains—tailings, foundation, and construction materials (low-permeability borrow and rock fill). Drained and undrained probabilistic shear strength distributions were derived from literature (Jefferies & Been, 2015; Christian & Baecher, 2011; Been & Li, 2009; Vick, 2002; Becker, Cavalcanti & Marques 2023) to reflect strength variability within each domain.

To develop controlled data sets, representative drained and undrained strength values were selected as base cases. Literature-reported ranges of CoV (Phoon & Kulhawy, 1999; Phoon & Ching, 2015; Ching, 2021, Uzielli, 2008), as well as relevant data from published statistical analysis of geotechnical parameters (Linero-Molina, Contreras & Dixon, 2021; Vernengo & Contreras, 2023) were used to fit both normal and log-normal distributions. Three CoV levels (minimum, mean and maximum) were applied to assess sensitivity to uncertainty.

In general, lognormal distributions were preferred, as they better represent the skewness and non-negativity of geotechnical strength parameters. To avoid unrealistic sampling in Monte Carlo/Latin Hypercube simulations, distributions were truncated at ±1E-08.

Due to limited data, a key limitation of this study is the use of univariate distributions in the stability models. The choice of distribution type was based on typical skewness patterns reported in the literature.

The shear strength distribution parameters used in this study are presented in Table 1.

TABLE 1

Summary of drained and undrained probabilistic shear strength distribution parameters.

Material	**Tailings**						**Clay foundation**	**Clay borrow**		**Rock fill**
	Drained		**Undrained**					**C**	**ϕ**	**ϕ**
Strength	°	°	USR	USR	USR	USR	kPa	kPa	°	°
Mean	24	33	0.25	0.3	0.35	0.4	100	23.7	22.4	38
COV1 %	10	5	15	15	15	15	18	19	4	6
COV2 %	12.5	12.5	25	25	25	25	25	26	8	10
COV3 %	15	20	35	35	35	35	38	33	13	13
Skewness	>1	>1	>1	>1	>1	>1	>1	<1	<1	>1
Distribution	Log-Normal							Normal		Log-normal

RESULTS

Figure 2 illustrates the significant scatter observed in PoF across the range of FoS values. For FoS within the intervals [1.3, 1.4] and [1.4, 1.5], the PoF varied by more than 6 and 8 orders of magnitude, respectively, depending on the inherent variability of the foundation and embankment materials. For comparison, SLM curves (categories I–IV) span 4 to 5 orders of magnitude over the same FoS ranges. However, it is important to note that while SLM curves reflect increasing levels of engineering confidence, the probabilistic results presented here are independent of any predefined level of engineering, as CoV ranges used in the study are within expected natural inherent variability. While it may be possible to reduce epistemic uncertainty using better-quality data, a limit is reached where further investigation would not significantly reduce CoV.

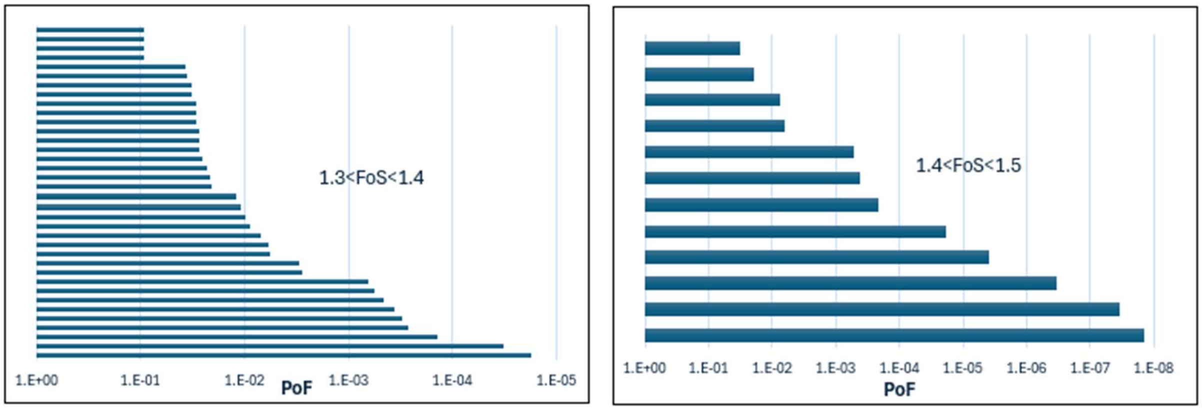

FIG 2 – PoF variability within FoS ranges.

A more detailed breakdown of PoF variability, focusing on the upstream raise case, is presented in Figure 3, showing PoF results for competent and weak foundations under varying mean values and CoV for undrained tailings strengths. In the competent foundation cases, variability in PoF increases with higher undrained strength, as the slip surface is governed primarily by the tailings. In contrast, results for the weak foundation do not display a clear trend, likely due to the interaction between the tailings and the weaker foundation materials influencing the slip surface.

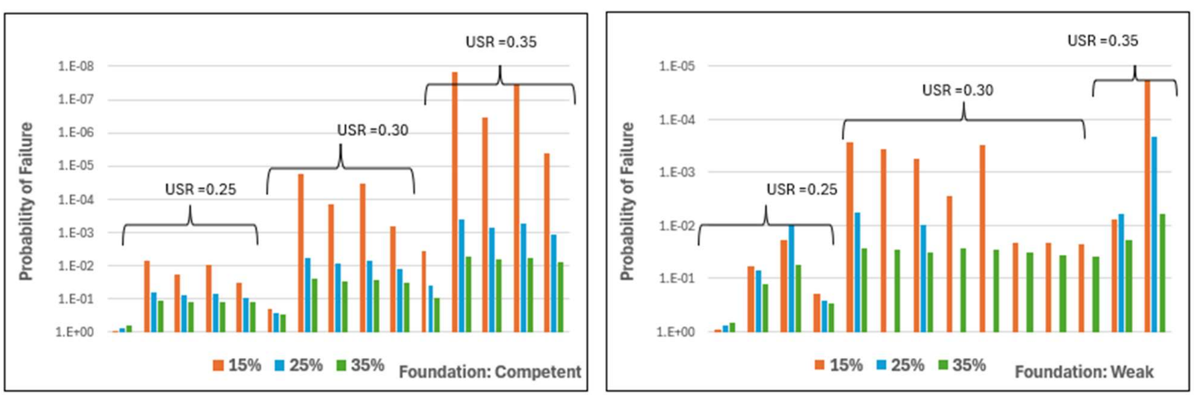

FIG 3 – PoF variability in upstream TSF cases considering competent and weak foundation conditions.

To enable direct comparison with the SLM framework, the probabilistic results were annualised as described earlier in the paper. Figure 4 presents these results across various operational cases, reflecting differing assumptions about the likelihood of triggering undrained loading conditions for each raise method as discussed earlier in the paper. A consistent trend is evident: most PoF values cluster around or above the SLM Category 1 curve—even in cases with relatively high CoV values and independent of raise method. This discrepancy suggests that case-specific probabilistic analyses, if interpreted without diligence, may yield PoF estimates that overstate the likelihood of failure, raising questions about how such results align with practical decision-making frameworks.

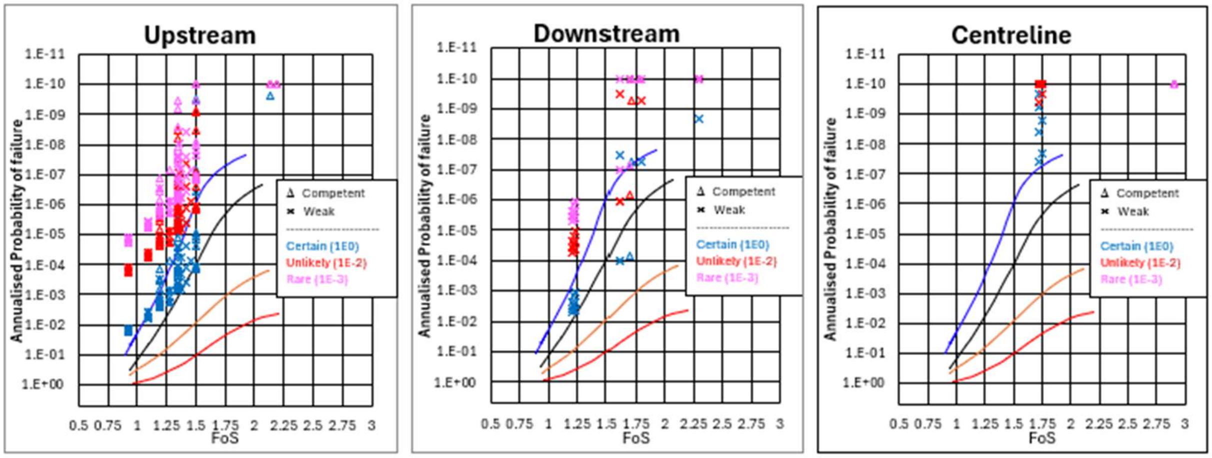

FIG 4 – Annualised PoF versus FoS varying likelihood of undrained triggers and foundation conditions.

For comparative purposes, a base case undrained strength ratio (USR) of 0.35 was selected for the tailings. The results, by foundation type and embankment raise construction method, considering that triggering an undrained response would have a likelihood of one in any given year, are shown in Figure 5. This plot offers a more granular view of PoF scatter and represents what might be encountered in a trade-off study. Interestingly, the results suggest that all raise construction types yield broadly similar PoF ranges under competent foundation conditions. Given that the critical slip surfaces analysed correspond to failure modes with severe consequences, the results imply that for this case, the downstream or centreline raise construction methods may not necessarily reduce PoF driven by undrained loading, provided foundation conditions are favourable.

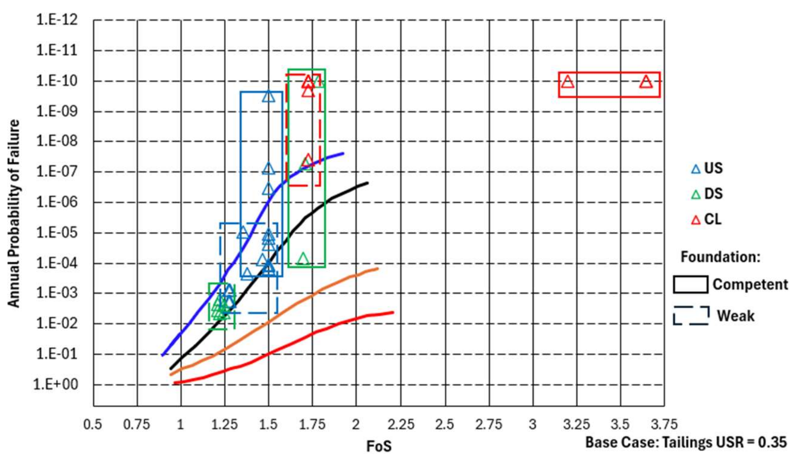

FIG 5 – Comparison of PoF between upstream, downstream and centreline methods of construction.

CONCLUSIONS

Selecting an approach to assess uncertainty in a geotechnical system is dependent on and constrained by regulatory, economic and technological factors.

This study highlights the need to exercise caution when using simplified methods such as SLM to assess TSF stability risk. Demonstrated misalignment between the FoS and PoF obtained from the SLM method can result in misrepresentation and overestimation of risk when used to establish a TSF's ALARP position. Where slope instability represents a major contributor to a TSF risk profile, costs to further investigate material variability would be justifiable. For many large-scale TSF operations, however, fully objective assessments remain challenging due to inherent limitations in data availability, precision, and accuracy.

Based on the findings of this study, the following conclusions are presented:

- The SLM method remains a useful screening tool in dam risk assessments; however, its applicability to TSFs is constrained by limited calibration, simplified treatment of foundation and loading conditions, and potential for misrepresentation of risk when deterministic FoS values are derived from conservative strength parameters.
- This study illustrates that PoF can vary by several orders of magnitude within narrow FoS bands, highlighting the significance of shear strength uncertainty particularly in undrained conditions. However, the application of probabilistic methods would require a critical view of the availability and potential variability of the geotechnical data.
- A practical way forward involves establishing bounded estimates of risk by identifying credible failure modes, defining realistic strength variability ranges through reasonable estimation of CoV, and applying a tiered investigation strategy. Literature-derived parameter distributions may offer meaningful input where site-specific data are limited. The aim is not to replace or downplay the importance of judgment; rather, to provide alternative framework for the rational application of geotechnical expertise.
- While conservatism is often applied as a safeguard, it can distort PoF estimates when probabilistic tools are used without adjusting input distributions accordingly. This may lead to either overstating risk or unjustified overdesign, both of which can undermine ALARP decision-making.
- Future work should aim to define practical guidelines or frameworks that identify minimum CoV targets that are achievable through standard testing and investigation methods, to help guide when a probabilistic analysis is meaningful and when it may be misleading.
- The authors are interested in collaboration using site-specific inputs to further refine a practical approach to quantifying TSF instability risks and informing ALARP assessments.

REFERENCES

ANCOLD, 2019 Guidelines on Tailings Dams, Planning, Design, Construction, Operation and Closure Addendum, dated July 2019.

Baecher, G, 1984. Geostatistics, Reliability and Risk Assessment in Geotechnical Engineering, *Geostatistics for Natural Resources Characterization*, Part II, pp 731–744.

Barneich, J, Majors, D, Moriwaki, Y, Kulkarni, R and Davidson, R, 1996. Application of reliability analysis in the environmental impact report (EIR) and design of a major dam project, Uncertainty '96, ASCE.

Becker, L D B, Cavalcanti, M d C R and Marques, A A M, 2023. Statistical Analysis of the Effective Friction Angle of Sand Tailings from Germano Dam, *Infrastructures*, 8:61. https://doi.org/10.3390/infrastructures8030061

Been, K and Li, A L, 2009. Soil Liquefaction and Paste Tailings, in *Paste 2009: Proceedings of the Twelfth International Seminar on Paste and Thickened Tailings* (eds: R Jewell, A B Fourie, S Barrera and J Wiertz), pp 281–290 (Australian Centre for Geomechanics: Perth). https://doi.org/10.36487/ACG_repo/963_32

Ching, J, 2021. State-of-the-art review of inherent variability and uncertainty in geotechnical properties and models, International Society of Soil Mechanics and Geotechnical Engineering (ISSMGE) – Technical Committee TC304 Engineering Practice of Risk Assessment and Management. https://doi.org/10.53243/R0001

Christian, J T and Baecher, G B, 2011. *Unresolved Problems in Geotechnical Risk and Reliability*, Geotechnical Special Publication.

Christian, J T, Ladd, C C and Baecher, G B, 1994. Reliability applied to slope stability analysis, *Journal of Geotechnical Engineering*, 120(12):2180–2207.

Duncan, J, Wright, S and Brandon, T, 2014. *Soil Strength and Slope Stability,* 2nd edn (John Wiley and Sons, Inc: Hoboken).

Harr, M E, 1977. *Mechanics of particulate media – a probabilistic approach* (McGraw-Hill Publishers: New York).

Jefferies, M and Been, K, 2015. *Soil Liquefaction: A Critical State Approach*, 2nd edn (CRC Press).

Linero-Molina, S, Contreras, L F and Dixon, J, 2021. Estimation of shear strength of very coarse mine waste, in *SSIM 2021: Proceedings of the Second International Slope Stability in Mining* (ed: P M Dight), pp 341–354 (Australian Centre for Geomechanics: Perth). https://doi.org/10.36487/ACG_repo/2135_21

Phoon, K-K and Ching, J, 2015. *Risk and reliability in geotechnical engineering* (CRC Press: Taylor and Francis Group).

Phoon, K-K and Kulhawy, F, 1999. Evaluation of geotechnical property variability, *Canadian Geotechnical Journal*, Nov.

Silva, F, Lambe, T W and Marr, W A, 2008. Probability and risk of slope failure, *Journal of Geotechnical and Geoenvironmental Engineering*, Dec:1691–1699.

Stark, T D, Moya, L J and Lin, J, 2022. Rates and Causes of Tailings Dam Failures, *Advances in Civil Engineering*, Hindawi.

Uzielli, M, 2008. Statistical analysis of geotechnical data, in *Reliability-based design in geotechnical engineering: Computations and applications* (ed: K K Phoon), (CRC Press). https://doi.org/10.1201/9780203883198.ch10

Vernengo Lezica, I, Contreras, L F, 2023.Application of Bayesian methods to the estimation of shear strength parameters for slope stability analysis of a TSF , *Mine Waste and Tailings Conference 2023*, Brisbane, 13-14 July.

Vick, S G, 2002. *Planning, Design and Analysis of Tailings Dams*, 2nd edn (BiTech Publishers Ltd).