PROBABILITY OF DAM SLOPE FAILURE AND DECISION-MAKING

Jiri Herza[1] and Hugo Fellows-Smith[2]

[1] HATS Consulting, [2] University of Western Australia

https://doi.org/10.56295/AGJ60410

ABSTRACT

Contemporary approaches to estimating slope failure probabilities of dams combine two main frameworks. One employs numerical models that represent input variability with statistical distributions. The other relies on empirical relationships and expert judgement to account for unquantifiable factors and limited knowledge. Whatever the analytical sophistication, any probability assigned to the one-off event of slope failure eventually incorporates a subjective degree of belief.

Despite its subjectivity, any framework used for safety-critical decisions should be internally consistent and logically coherent, so resulting decisions remain ethically and legally defensible. With this principle in mind, this paper examines the method introduced by Silva et al. (2008), a framework designed to inform dam-safety decisions and widely adopted in Australia and overseas. Using formal logic as the primary lens, the paper evaluates the coherence of the Annual Probability of Failure – Factor of Safety – Project Category approach without assessing data availability or judging practical convenience.

Our analysis reveals three key points. First, the anchor points that define the curves are not supported by a probabilistic model or empirical observations. Second, the fixed ordering of Project-Category curves does not fully reflect the key premise that more rigorous engineering reduces uncertainty. Third, the progressive flattening of the curves could discourage practitioners from adopting proven risk-control measures.

To strengthen its logical footing, this paper restates the framework as a formal theorem that makes every assumption explicit and outlines adjustments that resolve the inconsistencies. Well-documented cases where failures occurred despite apparently adequate Factors of Safety motivate an alternative probability estimate based on the likelihood of a critical engineering error. The paper concludes with a streamlined procedure for addressing foreseeable slope-instability risks within the Australian context.

1. MOTIVATION

The greatest tragedies are often labelled as "black swan" events, a term that allows individuals to evade admitting that earlier action could have been taken.

While some adverse events have been extensively studied, modelled, and assessed, others appear entirely unprecedented, shocking observers when they occur. This perception stems from the vast number of possible adverse scenarios, our incomplete understanding of underlying processes, and the limited resources available for their study.

The Swiss National Bank's decision on 15 January 2015 to scrap its CHF 1.20 per EUR cap caused the largest one-day move in a major currency since the 1970s. Foreign-exchange broker FXCM lost USD 225 million in fifteen seconds and called the event a "40-sigma" shock (Euromoney, 2015), implying a probability below 10^{-800}. Yet the scenario was never impossible: the SNB retained unilateral control over the cap and had publicly hinted at its removal months earlier (WSJ, 2014), while similar policy reversals by other central banks were well documented (Askari & Mirakhor, 2015). FXCM's loss arose because its risk models ignored a policy-driven discontinuity of that scale. The episode shows how omitting a plausible mechanism can render probabilistic estimates meaningless, a failure mode that becomes critical where the consequences are severe, as with dam failures.

Every dam failure is unique because the exact combination of structural conditions and triggers cannot be replicated. Although sabotage, warfare, and deliberate removal can cause failure, dams are designed and operated for stability and longevity, so unintended failures arise only under unforeseen or unaccounted-for conditions.

The 2019 Brumadinho tailings-dam disaster in Brazil exemplifies this problem. The owner, Vale S.A., reported that the dam contained 94 piezometers and 41 water-level indicators, with data collected and analysed regularly (Vale, 2019). Routine inspections revealed no deterioration, and Vale stated that "*the dam had a Safety Factor in accordance with world best practices and was above the Brazilian Standard reference point*." Two "Stability Condition Statements" issued by TÜV SÜD do Brasil in 2018 (TÜV SÜD, 2018) likewise attested to the dam's physical and hydraulic safety; the lowest pre-failure factor of safety reported was 1.09.

The case highlights a critical question: what does a pre-failure factor of safety actually signify when failure still occurs despite apparent compliance with accepted standards?

Faced with incomplete knowledge, engineers turn to risk assessment to evaluate possible outcomes, represent uncertainty, and judge the plausibility of extreme events such as dam failures. Yet the FXCM and Brumadinho examples show that a highly regulated environment, complete with multiple checks and balances, can fail abruptly when a governing failure mechanism is missing from the risk catalogue.

Acknowledging that dams can fail in many ways, this paper concentrates on embankment failures driven by slope instability. Specifically, it examines how the pre-failure factor of safety (*FoS*) is interpreted, how relevant it remains after failure, and how it informs decision-making.

Maintaining an adequate *FoS* is a cornerstone of engineering practice and is embedded in guidelines, standards, and legislation. The ANCOLD Guidelines on Tailings Dams (ANCOLD, 2019) provide an Australian example, recommending minimum *FoS* values that accommodate modelling limitations, material variability, and other uncertainties.

FoS is often converted to a probability of failure for slope instability. Silva et al. (2008) extended Lambe's (1985) ideas by creating an empirical relationship that links slope stability *FoS* and project category (*PC*) to annual probability of slope failure (P_f). Their framework, based on expert elicitation, blends calculated *FoS* with engineering diligence. Recognising that higher *FoS* values do not always imply lower P_f, the method has gained wide acceptance and is cited in the ANCOLD Guidelines on Risk Assessment (ANCOLD, 2020).

Other probability-estimation approaches derive from reliability theory, which models component interactions within a formal probabilistic framework (e.g. Whitman, 1984). Although reliability methods are well developed in many fields and have been studied for geotechnical structures (Christian et al., 1992; Dunken, 2000; Hartford & Baecher, 2004 and Xiao et al., 2020), uptake for embankment dams remains modest. Key obstacles include limited high-quality failure data, difficulties in producing robust phenomenological models, and lingering gaps in fundamental soil-mechanics knowledge (Mitchell & Soga, 2005). Consequently, reliability-based estimates for slope instability still rely, at least in part, on expert judgement (Hicks & Li, 2021).

Given its broad use in Australia, this paper focuses on the Silva et al. (2008) framework that the authors offered for "*risk-based decision making for situations involving slope failures.*" Under Australian law, safety decisions must be defensible, reasonable, and based on sound rationale. We therefore test the P_f-*PC*-*FoS* relationship proposed by Silva et al. (2008) for logical consistency, without judging the quality of its data, its predictive power or practical utility. In this paper we:

- Construct an explicit regression function that represents the relationship inferred by Silva et al. and test it for internal coherence
- Examine the method presented by Silva et al. (2008) through the lens of formal logic
- Outline modifications that resolve identified consistency issues
- Propose an alternative probability model based on the likelihood of critical engineering error
- Present a practical method that addresses slope-instability concerns within the Australian legal framework without requiring a quantitative failure-probability estimate

2. PHILOSOPHICAL BASIS OF SUBJECTIVE PROBABILITY

Though we can construct a formal logical system to quantify a subject's degree of belief, which we may call probability, such probability has no physical meaning.

Recognising the limitation of our knowledge, data and tools for dam risk assessment, ANCOLD (2022) defines probability as: "*A measure of the degree of confidence in a prediction, as dictated by the evidence, concerning the nature of an uncertain quantity or the occurrence of an uncertain future event. It is an estimate of the likelihood of the magnitude of the uncertain quantity, or the likelihood of the occurrence of the uncertain future event*."

Furthermore, ANCOLD (2022) clarifies: "*This type of probability is not a property of the dam, but a reflection of the best understanding of the analysis team, given the available knowledge and data concerning the question at issue*."

This view echoes Laplace (1774), who described probability as a measure of our ignorance, implying that with complete knowledge, probability statements would collapse into tautology. Chapter 6 will revisit this idea when we propose an alternative estimate of slope-failure probability based on the chance of critical engineering error.

When assessing the probability of dam slope failure through *FoS* estimation, we must recognise that the estimated *FoS* is not a physical feature of the dam. Instead, it is a computed (expected) ratio of resisting and driving forces calculated using methods and models that significantly simplify the very complex processes (e.g. Mitchell and Soga, 2005). Despite its name, the *FoS* values alone do not explicitly define safety, as explained in Silva et al. (2008). The same applies to the *PCs*, which are judgement-based indicators of uncertainty in input data and their analysis, not inherent properties of the project.

At this point, it is worth clarifying a common misconception that the method presented in Silva et al. (2008) was based on best fitting of observed annual failure frequencies of real-like projects separated into four categories. While the projects analysed by Silva's team were real, the P_f values presented Silva's team degree of belief in the notion of slope failures informed by their own calculations of *PC* and *FoS* variables. In Silva et al.'s (2008) own words "*the failure probabilities reflect their judgement of the relative degree of safety of the various earth structures.*"

Due to their subjective nature, the P_f-*PC*-*FoS* relation is meaningful only within the context established by Silva's team. Nonetheless, if it is to guide rational decisions, the underlying reasoning must be logically sound. For this paper we adopt the general conception of logic as the discipline of rational thought that ensures that the beliefs and their analytical products are not self-contradictory (Ramsey, 1926).

3. THE SILVA METHOD

Risk assessment holds no greater value than the utility of its outcomes, for its worth is measured not in the precision of probabilities and consequences, but in the insight, it affords to action.

Silva et al. (2008) adopted, refined and generalised the relationship between *PC, FoS* and P_f previously used to assess risks of slope failure of the coastal Amuay cliffside in Venezuela (Lambe, 1985). Lambe (1985) presented this relationship only in a graphical form and explained that "*We prepared Figure 21* [reproduced in Figure 1 – left] *as a numerical expression of our judgement – we had little data to prepare these plots*."

Silva et al. (2008) explained that his team, including Dr. Lambe and Dr. Marr together with their associates also developed their plot (Figure 1: - right) through quantification of the team's expert judgement.

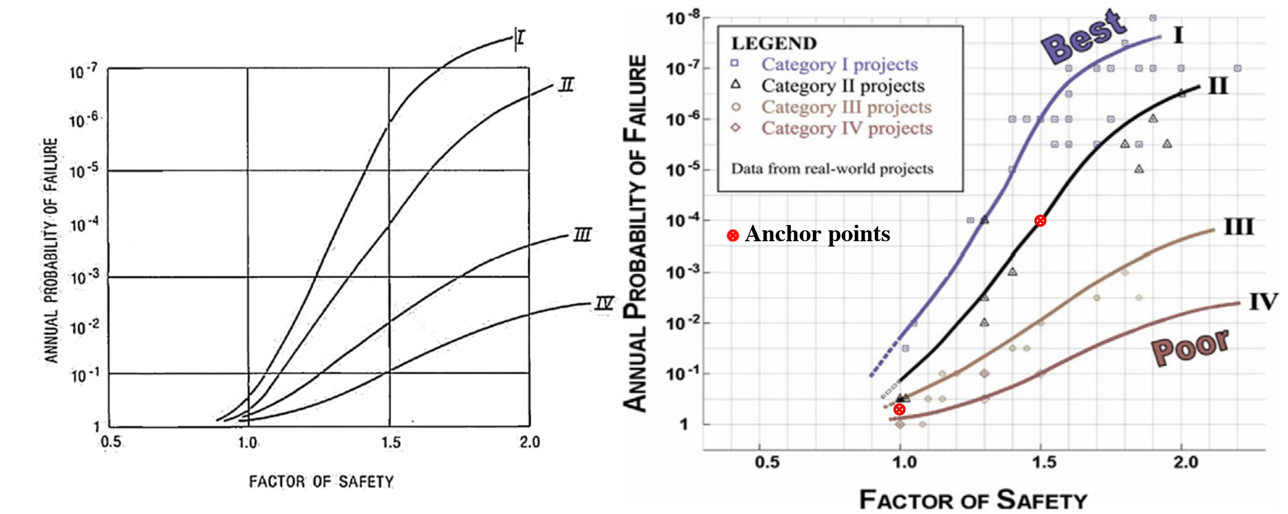

Figure 1: Annual Probability of Failure vs Factors of Safety, Figure 21 from Lambe (1985) – left, Figure 1 from Silva et al. (2008) - right

Silva's team analysed previously completed projects that included zoned and homogeneous earth dams, natural and cut slopes, and earth retaining structures. Based on their judgement of the level of engineering applied in the facilities design, construction and operation, the projects were assigned *PCs* from I to IV. Facilities designed, built, and operated with state-of-the-practice engineering were assigned Project Category I (PC_1); projects using standard engineering practice were assigned Category II (PC_2); projects with no site-specific design and substandard construction or operation were referred to as Category III (PC_3) Projects; and Category IV Projects (PC_4) were constructed with little to no engineering. Silva et al. (2008) provided a tabulated framework, similar to the one presented in Lambe (1985), for deriving the *PC* value and a weighted average of five partial factors.

Silva's team then performed stability analyses to determine the *FoS* values for the previously completed projects. Through an elicitation process, they then estimated P_f for each project by comparing the project with other projects and with two pre-defined anchor points (shown in Figure 1:, right).

The first anchor point was based on the work by Vick (1994) who estimated that the *FoS* = 1.0 yielded a probability of slope failure of 0.5, assuming a log-normal distribution of the FoS uncertainty. Silva et al. (2008) stated that this anchor point was assigned to PC_3 assets. The other anchor point (FoS = 1.5, $P_f = 1\times10^{-4}$) was assigned to PC_2 assets based on historical probabilities.

After agreeing on the position of the discrete projects in the FoS-P_f space, Silva et al. (2008) drew lines to fit the data points for each PC. These curves group projects with similar uncertainty and reflect the view that a larger FoS does not necessarily imply a smaller P_f, because its effect can be negated by the presence of larger uncertainties.

Silva et al. (2008) stated that the general shape of their four curves agrees with theoretical probability distributions without providing formal definitions, explicit assumptions or the probability distribution type. For practical use, Silva et al. (2008) encouraged the reader to interpolate between the regression lines to estimate the annual probability of slope failure for specific earth structure based on the estimated FoS and the PC value.

As the definitions and the key assumptions are important for finding a suitable probability framework and our later discussion, we have deduced these from Silva et al. (2008) and present them in Table 1.

Table 1 Summary of key assumptions inferred from Silva et al. (2008)

#	Definition/Assumption	Logical notation
1.	$P_f(FoS, PC)$ maps reported FoS and PC to the probability of failure in frequentist sense.	$P_f(FoS, PC) := \mathbb{P}[\text{fail} \mid FoS, PC]$
2.	PC is a proxy for the uncertainty in FoS estimates.	FoS is a random variable with standard deviation $\sigma_{FoS}(PC)$
3.	Improved engineering reduces uncertainty. Using Silva et al. (2008)'s convention, high PC represent higher uncertainty in FoS estimate.	$PC_2 > PC_1 \Rightarrow \sigma_{FoS}(PC_2) > \sigma_{FoS}(PC_1)$
4.	There is a PC-specific relationship between P_f, and FoS such that P_f for any given $FoS > 0$ is always greater for higher PC.	$PC_2 > PC_1 \Rightarrow P_f(FoS, PC_2) > P_f(FoS, PC_1)$
5.	Increasing FoS for a given slope reduces its probability of failure.	$FoS_2 > FoS_1 \Rightarrow P_f(FoS_2, PC) < P_f(FoS_1, PC)$
6.	P_f asymptotically approaches 1 when FoS approaches 0 for all PCs.	$\forall PC, \lim_{FoS\to 0} P_f(FoS, PC) = 1$

4. MATCHING REGRESSION MODEL

A dam failure caused by slope instability is a binary outcome—a singular moment in which the complexity of countless interactions resolves into either collapse or stability, with no in-betweens.

4.1. LOGISTIC MODEL

In this chapter, we address the missing formula for the theoretical probability distributions by examining the characteristic S-like shape of the regression lines and constructing a regression function that explicitly captures the essence of the P_f-PC-FoS relationship. As a dam slope failure is a binary event, a logistic regression, being a statistical model designed to predict binary outcomes based on one or more predictor variables, initially appeared to be a reasonable candidate for formalising the P_f-PC-FoS relationship.

At the core of the logistic model is the logistic function that models the probability of an outcome, in this case a dam slope failure, of the binary process:

$$P(y = 1|\boldsymbol{X}) = \frac{1}{1 + e^{-(a_0 + a_1 X_1 + a_2 X_2 + \cdots a_p X_p)}} \quad (1)$$

Where P(y=1|x) is probability that the outcome y is 1 (e.g., slope failure), x_1 , x_2 , …., x_p are predictor variables, a_o is the intercept term and a_1 , a_2 , …., a_p are coefficients that measure the effect of predictor variables on the outcome

This logistic function ensures that the predicted probability remains between 0 and 1 regardless of the values of the predictors and the typical S-shape logistic regression function plot is provided in Figure 2:.

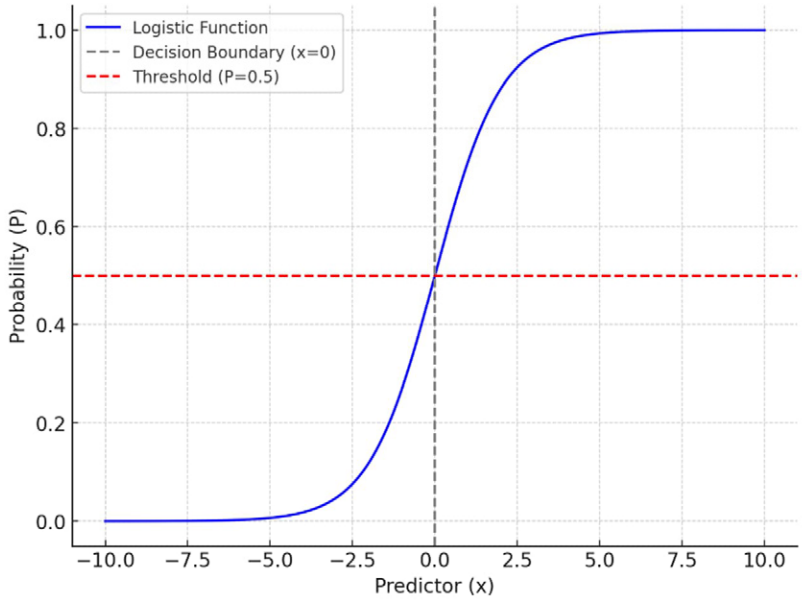

Figure 2: Typical logistic regression function

Logistic regression models perform optimally when the number of occurrences and non-occurrences of the binary outcome are similar (e.g. Chang, Dalpatadu, & Phanord, 2018). In the dataset and the symmetry around the decision boundary is apparent in Figure 2:. Given the extreme rarity of dam failure caused by slope instability, the logistic regression model in its natural form would not be suitable.

4.2. SUPER EXPONENTIAL FORM OF LOGISTIC MODEL

For modeling exceedingly rare events with binary outcomes, the logistic regression needs to be modified to respect the uneven nature of the outcomes and the expected form (shape) of the distribution curve in the corresponding scales.

As we attempt to formalise the approach presented by Silva et al. (2008) and achieve similar P_f-*FoS* relationship, our theorem is based on the assumptions listed in Table 1, except for assumption no. 3. We have to omit this assumption for the time being because it contradicts the shape of the regression lines in Silva et al. (2008), and we will revisit this problem later on.

In modifying the logistic function to correspond with the observed disproportionality of the outcomes (very rare events of dam slope failures), we first express the logistic probability as a logarithm of odds that the dam fails due to slope instability (P_f) as follows:

$$\text{logit}(P_f) = \ln\left(\frac{P_f}{1 - P_f}\right) = a_0 + a_1X_1 + a_2X_2 + \cdots a_pX_p \quad (2)$$

We then formalise the decay of the P_f with increasing *FoS*, acting as a predictor variable, as follows:

$$\text{logit}(P_f) = \ln\left(\frac{P_f}{1 - P_f}\right) = a_0 + a_1 \cdot FoS \quad (3)$$

The relationship between P_f and logit(P_f) is then expressed in terms of the logistic probability as follows:

$$P_f = \frac{e^{(a_0 + a_1 \cdot FoS)}}{1 + e^{(a_0 + a_1 \cdot FoS)}} \quad (4)$$

To formally capture the assumed non-linearity of the relationship between *FoS* and P_f we can denote the logistic term as logistic (*FoS*):

$$\text{logistic}(FoS) = \frac{e^{(b \cdot FoS - c)}}{1 + e^{(\text{b} \cdot FoS - c)}} \quad (5)$$

where "*b*" is a steepness coefficient similar to "a_1", and "*c*" is the threshold coefficient similar to "a_0" and this term presents the assumed sigmoid probability in logistic regression.

Because Silva's curves flatten at high *FoS*, we embed the logistic term inside an exponential decay so that P_f can approach a non-zero residual, with "a" being a scaling factor for "stretching" the probability values along the vertical axis as per the Silva et al. (2008) regression lines:

$$P_f = e^{-a \cdot logistic(FoS)} \tag{6}$$

By substituting the logistic probability back into equation (6), we obtain the following super exponential form of the logistic regression (SELR) function:

$$P_f = e^{-a\frac{e^{(b \cdot FoS - c)}}{1+e^{(b \cdot FoS - c)}}} \tag{7}$$

The parameter a governs the scaling of the probability decay, b determines the steepness of the transition, and c defines the threshold for the sigmoid-like progression along the *FoS* axis.

If *a, b* and *c* are non-zero real numbers, we can derive the limits of P_f when *FoS* approaches zero and infinity as follows:

$$\lim_{FoS \to \infty} e^{-a\frac{e^{(b \cdot FoS - c)}}{1+e^{(b \cdot FoS - c)}}} = e^{-a} \tag{8}$$

$$\lim_{FoS \to 0} e^{-a\frac{e^{(b \cdot FoS - c)}}{1+e^{(b \cdot FoS - c)}}} = e^{-a\frac{e^{-c}}{1+e^{-c}}} \tag{9}$$

The expression e^{-a} represents the residual risk of failure that resist all our efforts to prevent dam failures, while the second expression represents the remnant uncertainty on the other side of the spectrum where *FoS* is approaching zero and the probability is expected to approach one. While the characteristics of the SELR are consistent with regression lines presented in Silva et al. (2008), the parameters *a*, *b* and *c* lack empirical justification, and the model remains firmly within the realm of subjective probabilities and valid only for the key assumptions provided earlier.

4.3. MODEL PARAMETRISATION

The presented SELR form has a single predictor variable (*FoS*), which limits its generalisation and ability to represent the complex interplay of factors influencing slope stability. We can address this issue by adding predictors other than the *FoS*, such as embankment and foundation types, to better capture the multidimensional nature of slope stability. Introducing additional predictors may, however, overcomplicate the model without achieving more useful outcomes because the importance of potentially infinite candidates for additional predictors is not known, and the end calibration remains limited by the inherently small failure sample. Alternatively, we can define parameters *a, b* and *c* as functions of other parameters and factors.

We follow the latter approach and incorporate the *PC* variable to obtain regression curves similar to the regression lines presented in Figure 1: (right).

First, we express the variables *a, b* and *c* as linear functions of *PC* as follows:

$$a = m_a + k_a\, PC,\ b = m_b + k_b\, PC,\ c = m_c + k_c\, P\, C \tag{10}$$

where m_i and k_i are independent parameters.

By substituting the independent parameters, *a, b* and *c* in equation (7) with the above linear equations, we obtain the SELR function in this extended form:

$$P_f = e^{-(m_a + k_a\, PC)\frac{e^{[(m_b + k_b\, PC) \cdot FoS - m_c + k_c\, PC]}}{1+e^{[(m_b + k_b\, PC) \cdot FoS - m_c + k_c\, PC]}}} \tag{11}$$

The resulting SELR curves with the following set of m_i and k_i, superimposed by the regression lines drawn in Silva et al. (2008) are provided in Figure 3:.

$$m_a = 22.0\,, k_a = -4.0\,, m_b = 5.1\,, k_b = -0.3\,, m_c = 6.2, k_c = 0.0 \tag{12}$$

While the parameters *a*, *b* and *c* could be further optimised using more complex relations with the project category variable to "better" match the hand-drawn lines presented in Silva et al. (2008), we leave this to the reader. Our aim is to demonstrate that a mathematically consistent formulation can reproduce the qualitative features presented in Silva et al. (2008).

We will, however, revisit the consistency of the assumptions in Table 1 and logic behind the shape of the regression lines from Silva et al. (2008) in the next chapter.

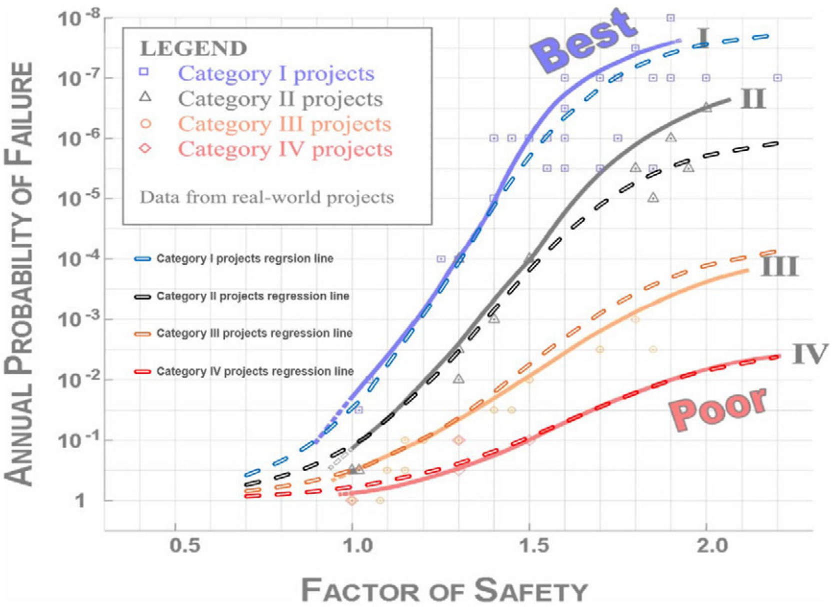

Figure 3: Comparison of regression lines from Silva et al. (2008) with super exponential form of the logistic regression

5. REVIEW OF CONSISTENCY AND COHERENCE

While the laws of logic and mathematics that describe the real world apply to our degrees of belief, they do not guarantee that our conclusions reflect reality—only the coherence of our thought.

5.1. OVERVIEW

We acknowledge that humans operate and thrive within internally inconsistent concepts such as love or art, where contradictions are expected, tolerated and even cherished. However, internal consistency and coherence are essential in probability frameworks used to inform safety-critical decisions.

The following examination of the method presented in Silva et al. (2008) focuses on its internal consistency and logical coherence without discussing the validity of the underlying data, the accuracy of the probability predictions, or the practical use of the method.

Silva et al. (2008) stated that their team used data from 75 projects to develop the relationships shown in Figure 1: – right without disclosing the underlying dataset or clarifying whether any of the analysed projects had actually experienced slope failure. Therefore, the *PC, FoS* and P_f values they assigned to the analysed projects can only be read from Fig.1 included in Silva et al. (2008).

5.2. ASSIGNMENT OF PROJECT CATEGORIES

Silva et al. (2008) stated that the *PC* should be calculated as a weighted average of five partial values taken from their Table 1. Yet, Silva's team assigned integer *PC* values to all their projects. As it is extremely unlikely (~3.4×10^{-53}) that weighted averages of five partial values for 75 different projects all happened to be integers, we consider that Silva's team likely rounded the *PC* values to the nearest whole number or directly assigned an integer *PC* value for each project. This could explain the broad spread of P_f values within the same category and the partial overlap of data points with different *PCs*. For example, a $P_f = 1{\times}10^{-7}$ was assigned to at least seven PC_1 projects with *FoS* values ranging from 1.6 to 2.2, while $P_f(FoS = 1.3) = 1{\times}10^{-4}$ appears in both PC_1 and PC_2 projects.

5.3. LINEAR VS. GEOMETRIC MEAN

Another notable pattern in Figure 1: – right is the plotting of P_f values at either major or minor gridlines denoting whole orders of magnitude, and geometric means of the closest orders, respectively. Silva et al. (2008) stated that a point ($P_f(FoS = 1| PC_3) = 0.5$) was used as one of the two anchor points for assigning subjective probabilities and fitting the regression lines. However, Figure 1: – right shows the regression line for PC_3 projects clearly intersecting $FoS = 1$ and the horizontal gridline presenting the geometric mean of $P_f = 0.1$ and 1.0. Because the geometric mean of 0.1 and 1.0 is not 0.5, the regression line for Category III Projects does not pass through the anchor point that Silva et al. (2008) stated was adopted. The pattern indicates the authors might have applied geometric rather than arithmetic averaging when plotting P_f values. In such case, it is possible that projects were assigned failure probabilities in whole or half orders of magnitude, with the latter appearing as geometric means in their Fig.1.

5.4. SHAPE OF REGRESSION LINES

When comparing Silva et al. (2008) to Lambe (1985) both provided in Figure 1, we find that their regression lines match almost perfectly for $FoS \geq 1.4$. Silva et al. (2008) declared this similarity when noting that their *Pf-PC-FoS* relationship differs from Lambe (1985) only for low *FoS* values, which the authors explained by stating that their analysis of structures "*on the verge of failure*" allowed them to refine curves near $FoS = 1$. Therefore, the resulting regression lines appear to match a pre-selected shape more closely than they match the distribution of P_f values assigned in the study, which would then explain the following observations:

- PC1: The regression line was drawn with 16 out of 25 data points falling below it
- PC2 and PC3: The regressions were drawn consistently above the data points, with each line having one data point above it
- PC4: The regression line is based on just four data points with $FoS \leq 1.5$

5.5. SELECTION OF ANCHOR POINTS

Silva et al. (2008) pre-selected an anchor point for Category II projects ($P_f(FoS = 1.5| PC_2) = 1\times10^{-4}$) "*based on historic probabilities*" while referring to Baecher et al. (1980), Whitman (1984), and Christian et al. (1992). An attentive reader would note that:

- Baecher (1980) proposed the annual probability of a dam failure as 1×10^{-4} as a reasonable default value based on a limited data set of dams constructed in the US between 1940 and 1984, and as he stated, his provisional estimate was pending the results of further research.
- Whitman (1984) stated that *"...it was suggested that a default value for annual risk of 10^{-4} be used in the absence of other information about the safety of a particular project*." Whitman further noted that "*This value of 10^{-4} was simply the average annual rate of failure for actual dams, as reported by several investigators*".
- Christian et al. (1992) did not provide an independent estimate and referred to Baecher et al. (1980) when stating that *"….the historical rate of failure of dams is about 1×10^{-4} per dam-year*".

Based on the above, it appears that $P_f(FoS=1.5, PC_2)=1\times10^{-4}$ used by Silva et al. (2008) as an anchor point was based on Silva's team agreement with prior judgement of others rather than empirical historical data. This is understandable given that estimating the frequency of dam failures is extremely difficult due to limited data regarding both the total numbers of dams and the name and nature of their failures as shown by Rana et al. (2022). Furthermore, as discussed in Herza et al. (2023), the reference class issue severely erodes the relevance of observed frequency of dam failures, even if such frequency is known, to specific assets.

5.6. COMPARISON WITH LOG-NORMAL DISTRIBUTION

Silva et al. (2008) claimed that the general shape of their regression lines agreed with the probability distributions proposed by Whitman (1984). However, a closer examination of the superimposed *Pf-PC-FoS* plots, presented in Figure 4, reveals significant differences. While the log-normal probability distribution function used by Whitman (1984) dictates the convex shape of the curves, Silva et al. (2008) presented arguments for sigmoid-like distribution shapes resulting in each *PC* pairs that consistently diverge from each other beyond the points of their intersection. For illustration, the P_f values for PC_1 projects at $FoS = 1.5$ provided by Silva et al. (2008) and inferred from Whitman (1984) differ 100 times.

The probability distribution functions inferred from Whitman (1984) intersect in a single point ($FoS = 1.0$, $P_f = 0.5$), and their order would invert for $FoS < 1$, because Whitman assumed an impartial distribution of FoS values. In contrast Silva et al. (2008) maintain relative order of project categories on both sides of $FoS = 1.0$. This feature implies either a reversal in uncertainty or a significant bias in *FoS* estimates that the Silva's team incorporated in their analyses.

We acknowledge that Silva et al. (2008) presented their P_f values as annual probabilities of slope failure while Whitman (1984) discussed atemporal probabilities of slope failure based on the principles of reliability theory. Although these two approaches may not be directly comparable, to our knowledge there is no published transformation method that would explain the vast differences between the P_f-*PC-FoS* shown in Figure 4.

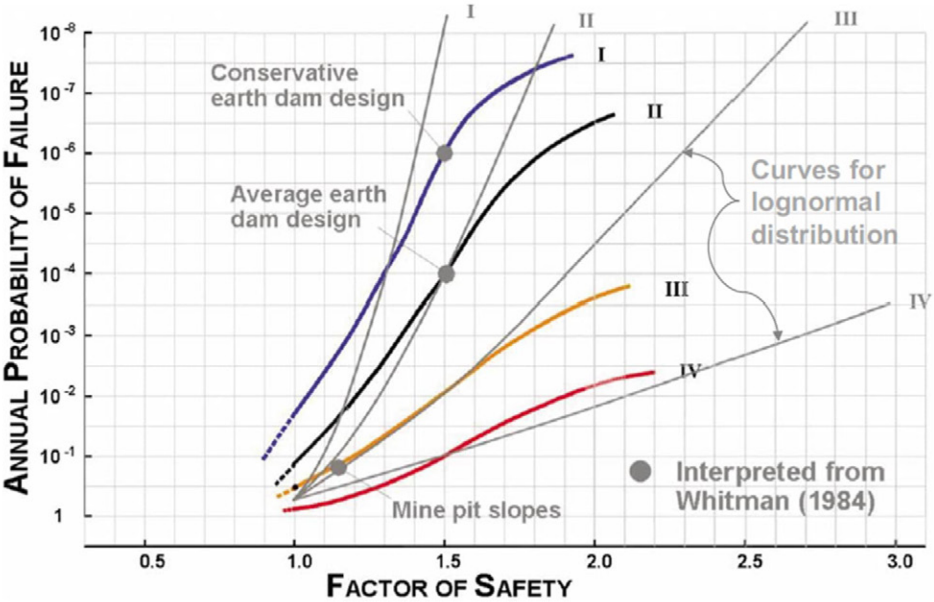

Figure 4: FoS versus Pf, with lognormal distribution and data from Whitman 1984 (adopted from Silva et al., 2008)

5.7. FIXED ORDER OF REGRESSION LINES

By maintaining the relative order of the regression lines irrespective of *FoS* (Assumption no. 4 in Table 1), Silva's team expressed their lower confidence in the slope instability predictions made for Category I Projects compared to Category IV projects for $FoS \leq 1.0$. This, however, contradicts the key principle used by Silva et al. (2008) that lower *PC* values represent comparatively smaller *FoS* uncertainty (Assumption no. 3 in Table 1). Using the key assumptions presented earlier we can formally show this contradiction as follows:

$$PC_i > PC_j \Rightarrow \sigma_{FoS}(PC_i) > \sigma_{FoS}(PC_j) \Rightarrow P_f(FoS < 1|PC_i) > P_f(FoS < 1|PC_j) \quad (13)$$

Equation (13) expresses the reversal in the order of the regression lines when $FoS < 1$, as a mathematical consequence of *PCs* being taken an impartial measure of uncertainty. However, this reversal is contradicted by the monotonic ordering of the regression lines observed in Silva et al. (2008), and mathematically reproduced as:

$$P_f(FoS < 1|PC_i) < P_f(FoS < 1|PC_j) \quad (14)$$

Silva et al. (2008) clarified that the shape of the regression lines near FoS = 1: "*come from the consideration that engineers inherently apply more conservatism in their choice of assumptions and selection of parameters as the importance of the structure and the consequences of failure increase. This produces a bias in their analyses such that their computed safety factors at low values does not represent the actual expected safety factor value, but something higher.*"

However, Silva et al. (2008) also advised the reader to determine *FoS* objectively, when using their method, stating that: *"The strength determination corresponds to the best estimate of the strength acting in the field and not necessarily the average strength or a "conservative" value of strength."*

This contradiction is not limited to the regression lines; it also extends to the individual project points established by Silva's team. Let us take the following points from Figure 1 – right as an example: $P_f(FoS = 1.02| PC_1) = 3.2\times10^{-2}$ and $P_f(FoS = 1.05| PC_1) = 1.0\times10^{-2}$. If Silva's team followed their own procedure and used their best engineering judgement to estimate the *FoS* and computed the *PC* values as weighted averages of the five defined factors, then assigning P_f values significantly less than 0.5 to these projects cannot be logically justified on the basis of conservatism in their own *FoS* estimates. Conversely, departing from the stated procedure would undermine the internal logic of the method.

Although Silva et al. (2008) discussed the influence of engineering conservatism in relation to PC_1 projects, the shape of all regression lines and the placement of individual points near $FoS = 1$ shown in Figure 1 (right), indicates a wider bias. For example, at least one PC_3 and one PC_4 project with $FoS = 1.0$, estimated by Silva's team, were assigned $P_f = 1.0$. Silva's team also assigned $P_f = 1.0$ to one PC_3 project with $FoS = 1.08$. These assignments suggest that the team treated *FoS* values for PC_3 and PC_4 projects as somewhat optimistic outliers, yet viewed some *FoS* values for PC_1 and PC_2 projects as deliberately conservative, which indicates methodological inconsistency.

5.8. FLATTENING OF REGRESSION LINES

Silva et al. (2008) clarified that the flattening of the curves near FoS of two reflected the team's belief "*that increasing the safety factor well beyond the typical values used for earth structures provides little benefit with respect to the corresponding probability of failure.*" Silva et al. (2008) then explained that "*discontinuities, weak zones, wet zones, high or low permeability zones, and other features that can elude a geotechnical investigation control the level of safety for grossly overdesigned facilities.*"

Through the flattening of the regression curves, Silva's team appears to be accounting for uncertainties beyond those captured by "*the standard deviation of FoS calculated with four distinct levels of engineering*". While it is admissible to assume that adverse conditions can evade a geotechnical investigation, the flattening of the regression lines at different levels suggests that increasing *FoS* through physical modifications, would not materially reduce the subjective probability of failure. For example, it implies that $P_f\,(PC_4) > 1\times10^{-3}$, irrespective of the geometrical arrangement of the PC_4 slopes.

This implication is inconsistent with principles of slope stability. A flatter slope of an embankment results in a higher ratio of resisting to driving forces and should therefore be perceived of having a lower P_f compared to a steeper slope given that all conditions other than the geometry stays the same. Further discussion on the practical implication of regression lines' flattening at different levels is provided in Chapter 7.

5.9. POTENTIAL METHODOLOGICAL IMPROVEMENTS

To address the formal inconsistencies discussed above, the work by Silva et al. (2008) could be further enhanced by following methodological adjustments:

- Objectivity in Slope Stability Calculations: Slope stability calculations should be conducted objectively and consistently, independent of the assigned *PC*, to prevent bias in *FoS* estimation.
- Transparent *PC* Assignment: *PC* should be assigned using a clear, consistent, and transparent methodology that allows for non-integer values, ensuring that uncertainty is accurately represented, and without considering the *FoS* values or consequences of the slope failure.
- Independent Expert Elicitation Process: The expert elicitation process to derive slope failure probabilities should be internally consistent, incorporate stability assessment results, and remain independent of *PC* assignment to eliminate potential bias.
- Three-Dimensional Representation of Results: The results of the expert elicitation process should be combined with the *PC* assignment and stability assessment results in a 3D matrix or plot, with *PC*, *FoS*, and P_f plotted along perpendicular axes.
- Formal Regression Analysis: A formal regression method should be applied to quantify the relationship between *PC*, *FoS*, and P_f (if any exists), ensuring that the obtained data are interpreted logically and consistently.
- The process should be clearly documented including defining the key terms, disclosing data, listing the key assumptions and outlining the methods used in every step.

In Chapter 6, we develop a formal theorem that resolves these inconsistencies.

6. FORMAL IMPROVEMENTS AND ALTERNATIVE APPROACH

Improving what exists and creating afresh are both essential pathways to meaningful advancement.

6.1. RESOLVING FORMAL INCONSISTENCIES

Obtaining SELR curves that resemble the regression lines in Silva et al. (2008) required omitting Assumption No. 3 (uncertainty increases with increasing *PC*) while maintaining Assumption No. 4 (constant order of regression lines), which would otherwise lead to contradictions, proven earlier.

Because engineering quality is widely accepted to reduce epistemic uncertainty, formalised in Assumption No. 3, we have to address the contradiction by removing Assumption No. 4 and formulating the following condition:

$$P_f(FoS = 1.0) = 0.5, \forall PCi \tag{15}$$

By removing the constant ordering (Assumption 4) and enforcing $P_f(1) = 0.5$ for all PCs, we eliminate the contradiction established in (13)–(14) while retaining Assumption 3. To satisfy the above condition within the context of the SELR (equation 7), we need to solve the following:

$$e^{-a\frac{e^{(b-c)}}{1+e^{(b-c)}}} = \frac{1}{2} \tag{16}$$

which is equivalent to

$$-a\frac{e^{\ (b-c)}}{1+e^{(b-c)}} = \ln\frac{1}{2} \tag{17}$$

Since $e^{(b-c)} = \frac{e^b}{e^c}$, we can reorganise equation (17) for the scaling factor a to:

$$a = \ln(2)\frac{e^b + e^c}{e^b} \tag{18}$$

The removal of the constant ordering of the regression lines (Assumption No. 4), also allows us to address the issue of asymptotic flattening of the regression at different P_f values, discussed in Section 5.8. To address the possibility that adverse conditions, evading our attempts to detect them, may present a residual risk even for state-of-art facilities, we can incorporate the residual risk level in the SELR. For illustration, we may assign an arbitrary number (P_r), greater than 0, as the residual P_f irrespective of PC and impose this condition by defining the scaling factor a as:

$$a = \ -\ln(P_r) \tag{19}$$

We then express the logistic value required at $FoS = 1$ as:

$$r = \frac{\ln 0.5}{\ln P_r}, (0 < r < 0.5)\ , \tag{20}$$

and solve the sigmoid threshold c for any positive slope parameter $b > 0$

$$c = b - \ \ln\left(\frac{r}{1-r}\right) \tag{21}$$

The exponential term $\frac{e^{(b \cdot FoS - c)}}{1+e^{(b\cdot FoS-c)}}$ is now equal to r when $FoS = 1$ and the resulting SELR gains the following form:

$$P_f \ = e^{\ln P_r \frac{e^{[b\cdot FoS - b + \ln\left(\frac{r}{1-r}\right)]}}{1+e^{[b\cdot FoS - b + \ln\left(\frac{r}{1-r}\right)]}}} \tag{22}$$

Like we did before, we can now relate the slope parameter b to PC as a proxy of FoS uncertainty, inspired by Silva et al. (2008). If desired, the sigmoid asymmetry in the lin-log scale shown in Figure 1 – right, can then be achieved by relating the slope parameter b to both the PC and FoS variables.

Figure 5 presents an example of two SELR curves derived from equation (22) with slope parameter b expressed as a polynomial function of PC. For comparison we superimposed the curves by a semitransparent segment of Figure 1 (right).

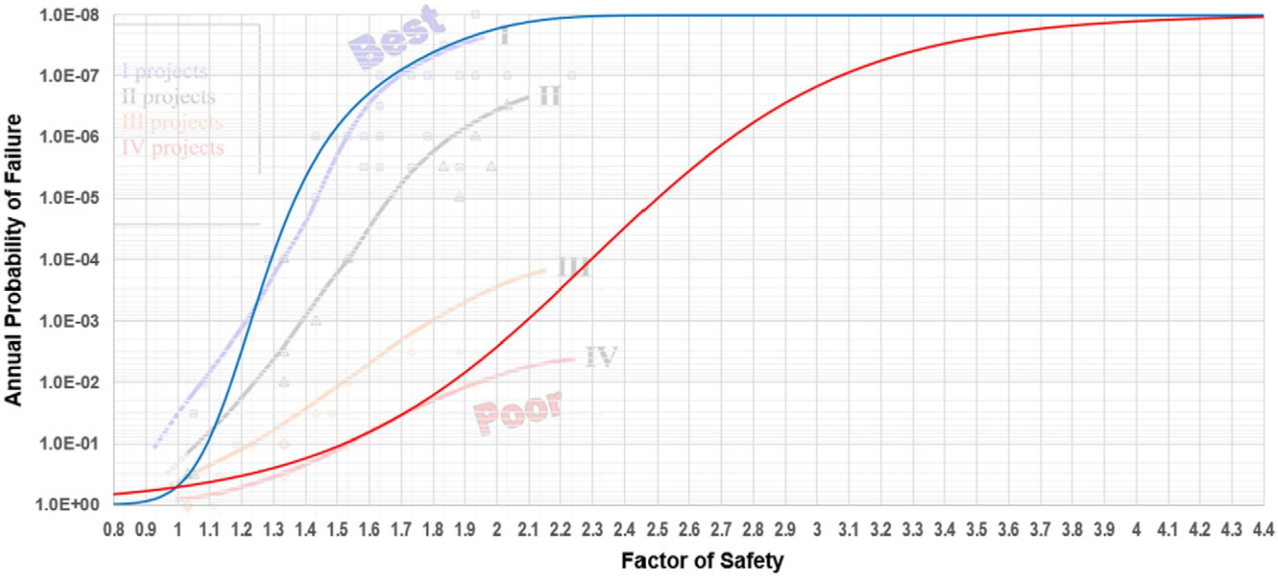

Figure 5: Examples of adjusted super exponential form of the logistic regression

Figure 5 now conforms to the assumptions that an impartial calculation of $FoS = 1$ should always return P_f value of 0.5 and that the probabilities of failure would asymptotically approach a common residual risk of failure, which we set at $1{\times}10^{-8}$ to be directly comparable with Silva et al. (2008). The residual probability of failure can, however, be set arbitrary close to zero, to respect the perception of the risk analyst. Figure 5 also supports the notion that structures with limited data but with very conservative FoS may be assigned similar subjective P_f values that are assigned to well-engineered structures with comparatively much smaller *FoS*.

Although the slope parameter *b* in equation (22) can be related to both FoS and the *PC* to achieve different sigmoid shape, there are more suitable methods of direct parameterisation to fit the desired probability shapes, such as the Metalog Distribution (MD), introduced by Thomas W. Keelin (2016) (Keelin, 2016). The generic MD form for estimating P_f as a function of FoS is as follows:

$$P_f(FoS) = \frac{1}{1 + e^{-M(FoS)}} \quad (23)$$

where M(FoS) is a Metalog function with a generic form of:

$$M(FoS) = a_1 + a_2 \ln(FoS) + a_3 FoS + a_4 FoS \ln(FoS) + a_5 (FoS)^2 + \cdots \quad (24)$$

Parameters a_i control the position and shape of the distribution function and can be related to *PC*. Additionally, the MD could be used as the argument of an external exponential function to achieve a desirable function shape in the lin-log scale, similar to the transformations applied to the logistic function earlier.

Given the subjective character of the P_f-*PC*-FoS relationship, the regression function can be parameterised by directly selecting coefficients that express this relationship, without resorting to anchor points, and thus sidesteps the inconsistencies that can accompany purely heuristic curve fitting.

6.2. ALTERNATIVE APPROACH

Irrespective of the method used to express the P_f-*PC*-FoS relationships, the probability of failure would be related to some form of statistical distribution of uncertainty associated with the FoS estimate. However, we observed that past slope failures of dams occurred under conditions that were not accounted for prior to the failure and hence could not be captured by the statistical distribution of uncertainty associated with the pre-failure FoS estimate.

To address this observation, we propose an alternative theorem for estimating slope failure probability that as the probability that the pre-failure assessment was significantly incorrect. Rather than directly relating the probability of failure to the pre-failure *FoS*, we instead link it to the error in the *FoS* estimate, emphasising the limitation of predictive analytical models we touched upon in the introduction section of this paper.

Using Silva et al. (2008) *PC*, the idealistic, "Category 0 Project" represents a fictional dam which is designed, built, operated and monitored such that the entirety of its composition and its surrounding environment is known, with no uncertainty regarding the dam's performance. For such a dam, the probability of failure is

$$P_f(FoS, 0) = \begin{matrix} 0 & FoS \geq 1 \\ 1 & \text{otherwise} \end{matrix} \quad (25)$$

Where, because $PC = 0$, the modeller's calculation FoS equals the actual (not modelled) ratio between the true stabilising and destabilising forces, true Factor of Safety ($tFoS$).

The calculated FoS reported by an analyst is, in practice, an estimation of this value derived from information that can only give inferences about the $tFoS$. For example, sampling the dam's material in several locations to calculate friction coefficients will give a probabilistic understanding of the true friction coefficients, and the way in which these values are mathematically combined assumes a degree of understanding about the mechanisms of failure. This means that the calculated *FoS* is wrong with respect to *tFoS* according to some distribution $Error_{PC}$, which we may represent as:

$$\frac{FoS}{tFoS} = Error_{PC} \quad (26)$$

Error is measured proportionally rather than additively. If the distribution of $Error_{PC}$ depends only on *PC*, the model never supposes $tFoS$ could be negative and $\frac{FoS}{tFoS}$ is independent of FoS.

$$\frac{FoS}{Error_{PC}} = tFoS \quad (27)$$

$$tFoS \leq 1 \Rightarrow Error_{PC} \geq FoS \quad (28)$$

If $f_{PC}(x)$ is the probability density function of $Error_{PC}$, then

$$P_f(FoS, PC) = \int_{FoS}^{\infty} f_{PC}(t)dt \quad (29)$$

This framework can be directly applied to decision-making processes, as the marginal improvement in failure probability given a marginal improvement in FoS is:

$$\frac{d}{dx}P_f(x, PC) = -f_{PC}(x) \tag{30}$$

In selecting an appropriate distribution for the error, a strong starting point is to use the centralised lognormal distribution with median = 1, as this is the canonical distribution for proportional error of a measurement. In this case,

$$P_f(FoS, PC) = \int_x^{\infty} \frac{1}{t\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(\ln t)^2}{2\sigma^2}\right) dt = \Phi\left(\frac{-\ln x}{\sigma}\right) \tag{31}$$

where Φ is the cumulative density function (equation 32) of the standard normal distribution, and anchor points or regressions may be used to fix the degree of freedom to a function of PC.

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x/\sqrt{2}} \exp(-t^2)\, dt \tag{32}$$

By expressing the error function (*erf*) as:

$$erf\left(\frac{-\ln x}{\sigma}\right) = \frac{2}{\sqrt{\pi}} \int_0^{z} \exp(-t^2)\, dt \tag{33}$$

and using symmetry and the Gaussian integral $\int_0^{\infty} e^{-t^2}\, dt = \sqrt{\pi}/2$, equation (31) becomes:

$$P_f(FoS, PC) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} \exp(-t^2)\, dt = \frac{1}{2}\left[1 + erf\left(\frac{x}{\sqrt{2}}\right)\right] \tag{34}$$

As a result, any set of error distributions with median $= 1$ and variance determined monotonically by PC tending to 0 as $PC \to 0$ will have the desirable properties that:

$$P_f(FoS = 1, PC) = 0.5\ \forall PC \tag{35}$$

and

$$\lim_{PC\to 0} P_f(FoS, PC) = P_f(FoS, 0) = \begin{matrix} 0 & FoS \geq 1 \\ 1 & \text{otherwise} \end{matrix} \tag{36}$$

Figure 6 shows the failure probability curves from the lognormal error model for σ values of 0.25 (blue), 0.1 (red), and 0.02 (green), illustrating that reduced epistemic dispersion (lower σ = lower PC) steepens the transition and lowers P_f for $FoS > 1$.

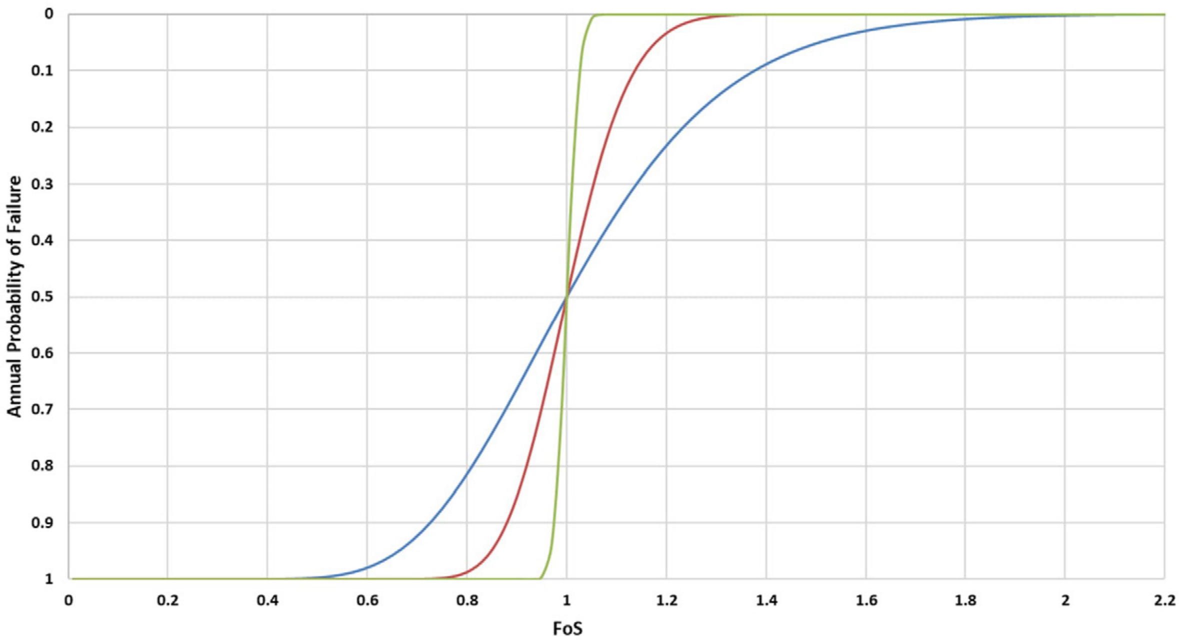

Figure 6: A plot of failure probability generated from a lognormal error distribution

If the method was to apply historical data to the risk assessment process, the notion of known unknowns versus unknown unknowns, and the fact that modes of failure in historical cases have been unexpected (e.g. Vale, 2019) may remind the analyst to consider the impact of higher PC on f_{PC}.

For example, consider partitioning $f_{PC} = \sum_i \chi_i f^i$ such that $\sum_i \chi_i = 1$, where each f^i is the probability density function for failure given the dam fails through a particular failure mechanism, or set of mechanisms, i; and χ_i is the observed historical proportion, within failed dams of similar PC, of failure through that mechanism or set of mechanisms. If these are grouped by the degree to which they have been considered and mitigated, then the impact of an increasing PC can be seen to continuously migrate modes of failure from less considered to more considered.

Even a simple model of two categories, "completely unconsidered" and "completely prevented", can represent the phenomenon where dam failure is often a black swan event, and hence acknowledges the limited ability to reduce failure probability beyond a certain residual level through improving PC.

The method offers potential for practical application, but the current dearth of data on modellers' sentiments about failed dams prior to their failure, and the degree to which the realised mode of failure was previously considered by modellers, is a major obstacle to implementation. Further development of this novel approach will require establishing credible ways of linking PC to a suitable error distribution and demonstrating the method's robustness through synthetic case studies.

7. SLOPE STABILITY CONCERNS IN AUSTRALIAN CONTEXT

The words 'reasonably practicable' are ordinary words bearing their ordinary meaning……And the question whether a [risk] measure is or is not reasonably practicable is one which requires no more than the making of a value judgement in the light of all the facts. Slivak v Lurgi [2001] HCA 6, [55].

7.1. AUSTRALIAN CONTEXT

The Australian legal context requires duty holders to make decisions affecting safety of workers and others based on rational, defensible and reasonable methods and through rational, defensible and reasonable systems. Because Silva et al. (2008) proposed a "*method of risk-based decision making for situations involving slope failures",* and their method is widely used across Australia, this chapter discusses the context of risk-related decision making in Australia.

Before estimating probabilities to support a dam risk assessment, we must understand the purpose of the assessment and how the outcomes will be used by the duty holder. Answering these questions often reveals that estimating the probability of slope failure is not always the most appropriate approach, nor is it necessary to achieve the intended outcomes.

This is illustrated by the probability and risk assessment example included in Silva et al. (2008), which concluded that installing relief wells to decrease pore pressures near the toe of the assessed dam "*made a lot of sense*". While we may disagree with the cost benefit rationale for this decision, we concur with Silva et al. (2008) that addressing the foundation conditions at the dam's toe, where personnel reportedly *"sank into the ground up to their knees"* made a lot of sense. However, we question whether the quantitative risk assessment was needed to arrive at this conclusion; a question that applies to all other conclusions from the example in Silva et al. (2008).

If probability-based assessments of dam slope failure are used to inform risk mitigation decisions in Australia, all involved parties must carefully consider the Australian legal context and the owners' obligations under the WHS Acts supported by case law, which are consistent with the obligations under other mining, water and dams acts. Greg Smith, a lawyer specializing in Work Health and Safety, clarified this legal context in a presentation and Q&A session organized by Engineers Australia in August 2024 (Smith, 2024).

We recognise that simplified quantitative risk assessments, cost-benefit analyses and comparison of risk position against risk tolerability thresholds, such as those discussed in Silva et al. (2008), are valuable in many situations. However, we must be mindful of the extent to which focusing on these assessments might distract the duty holder from fulfilling their fundamental duty to identify and implement all reasonably practicable measures to eliminate or reduce risks to health and safety, as is often the case in tailings dam failures.

Without straying too far from the technical nature of this paper, we stress that Australian WHS law (e.g., the Work Health and Safety Act 2020 (WA), Section 18, paragraph e) demands that when determining what is reasonably practicable in ensuring health and safety, the cost of risk control measures, and whether the cost is grossly disproportionate to the risk, may only be considered after assessing the available ways to eliminate or minimise the risk. Furthermore, to our knowledge, there is no legal precedent in Australia where a risk tolerability threshold has been accepted as a defence in a legal case concerning the safety of workers or others. Therefore, dam owners in Australia should exercise extreme caution before choosing not to look for risk reduction measures based on a risk assessment concluding that the risks fall below a specific tolerability or acceptability threshold, or that the dam is "*Adequately Safe*" as suggested in Fig. 3 of Silva et al. (2008).

7.2. CHALLENGES IN APPLYING PROBABILISTIC ESTIMATES IN AUSTRALIAN CONTEXT

It is worth noting that one of the primary reasons for introducing risk assessment into dam safety was to address areas where deterministic, standards-based methods were not available, such as piping phenomena or the reliability of spillway gates (ANCOLD, 2003). Standards-based, deterministic methods and their associated threshold values (e.g., minimum recommended FoS) can be considered implicit forms of risk assessment. ANCOLD (2022) acknowledges this link between standards-based FoS estimates and probability-based assessments in the following statement: "*For most dams, if the FoS for the steady state seepage condition is ≥ 1.5 using reasonable but conservative assumptions, and the monitoring of settlement (and lateral movements if available) indicates there are no unusual movements, the likelihood of failure by slope instability can be taken as negligible*."

The implicit conservatism in these methods must be acknowledged and appropriately addressed when using deterministic, standards-based slope stability outcomes to estimate dam failure probabilities. This is challenging because the various levels and types of implied conservatism are never explicitly stated. For example, slope stability analyses for embankment tailings dams are typically performed for specific, discrete sets of conditions (e.g., phreatic pressures, tailings levels, material characteristics), without the analyst explicitly defining their conditional probabilities that would be required for estimates of annual probabilities of slope failure. Beyond the fundamental challenge of incorporating deterministic, standards-based outcomes into probabilistic estimates, ANCOLD (2022) highlights additional difficulties, including unknown or unrecognized features and behaviours (unknown unknowns) that are not captured in the analysis. Given the historically low frequency of slope instability-triggered dam failures, ANCOLD (2022) concludes that in relation to slope failure "*the use of formal probabilistic methods is typically not warranted, even in the most detailed risk analyses*."

Available evidence suggests that many slope-instability failures might have been avoided had current ANCOLD-aligned practices been followed and we emphasize that compliance with current industry practices is a fundamental duty of dam owners in Australia. This duty extends well beyond slope stability considerations. In this context, we view the *PC* classification as equivalent to a gap assessment against current industry practices, with a Category I Project meeting current industry practice, while a Category IV Project has substantial deficiencies.

These deficiencies can be resolved by addressing gaps in current practice related to the understanding of materials, geometry, and loading conditions that control slope stability. Alternatively, the uncertainties can be managed by adopting a solution that ensures stability, so far as is reasonably practicable, under all foreseeable conditions. In practice, this may involve flattening the slope in high PC projects. Such a solution may be appropriate at mines where legacy embankments and other structures have a seemingly sufficient FoS though supported by limited engineering knowledge (e.g., PC_3 and PC_4 structures), and where construction materials, equipment, and workforce are readily available.

In such circumstances, physically flattening the downstream slope, without having a substantially improved engineering knowledge, may still be a reasonably practicable risk reduction measure, provided the construction activities are appropriately planned and managed. For example, conventional slope flattening by pushing material from the upper portion of the slope downward reduces the overall slope angle without adding new surcharge, and thus lowers the risk of failure once construction is complete. Buttressing slopes is another conventional method of reducing the risk of slope instability, particularly where deep foundations do not govern slope stability, as is often the case with in-pit structures.

The regression lines from Silva et al. (2008) suggest that such works would have virtually no effect on the subjective annual probability of slope failure, which could discourage the dam owner from implementing effective and well-understood risk reduction measures. Instead, the owner may opt to conduct additional investigations and analyses aimed at improving the PC, which, according to Silva et al. (2008), could result in a significant reduction in P_f.

7.3. RATIONAL METHOD FOR ADDRESSING SLOPE STABILITY CONCERNS IN AUSTRALIA

If, after meeting current industry practices, an owner seeks to further strengthen their embankment slope safety, perhaps as part of compiling a safety case demonstrating that all reasonably practicable risk controls are in place (Herza et al., 2022), we propose the following process. This process is inspired by the method outlined in Appendix B of ICOLD Bulletin 194 (ICOLD, 2023), which advocates for a rational approach to uncertainty and provides guidance for decision-making when existing tailings facilities do not meet the recommended FoS values. Since we believe this fundamental approach is applicable regardless of specific FoS values, we propose the following steps:

- Determine the worst foreseeable conditions, objectively informed by evidence, which could lead to slope failure resulting in an uncontrolled release of the dam content.
- Determine physically possible measures that would increase the FoS for any condition or combination thereof.
- Assess, from the owner's perspective, which of the identified measures are not reasonably practicable to implement. This assessment may include a probability analysis of specific phenomena (flood, seismic event etc.)
- Implement all reasonably practicable measures to reduce the risk of slope failure.

We acknowledge that owners have limited resources, which may prevent them from implementing all reasonably practicable risk controls across their entire portfolio simultaneously. In this context, risk assessment, including probability estimation, plays an important role in enabling comparative analysis (e.g., portfolio risk assessments) and informing prioritisation of risk control measures.

If a quantitative risk assessment is conducted for comparison or compliance purposes, an owner may choose a method that prioritises internal consistency and logical coherence over purely heuristic methods based on subjective interpretation of probability estimates subjectively made by the interpreters. For such circumstances, this paper offers improvements to the Silva et al. (2008) method together with an alternative methods for estimating probabilities of slope failure.

In Australia, however, these methods must not be the sole basis for decisions over risk-control measures for a particular dam under specific conditions, because relying on them alone may fail to meet ethical expectations or regulatory and legal scrutiny.

8. CONCLUSIONS

Progress comes from adding to what serves us well, subtracting what doesn't, and advancing with ever clearer reasoning.

While we acknowledge the widespread occurrence and benefits of inconsistent frameworks in our society, we consider that internal consistency and coherence are essential in probability frameworks used to inform safety-critical decisions in Australia.

Silva and his team successfully achieved their objective of making quantitative risk assessment more accessible to geotechnical engineers. Their methodology is now widely recognised and successfully applied as a contemporary approach to evaluating the probability of slope instability, both in Australia and internationally.

Analysing the extensive work by Silva et al. (2008) we have shown that applying a formal regression function to express subjective probabilities is not only possible but also desirable. It avoids inconsistencies that may otherwise be introduced when probability relationships are heuristically established from data that were previously approximated by the same authors.

Through further analysis of the method presented in Silva et al. (2008), we highlighted the importance of logical coherence and proposed steps that could resolve the formal contradictions associated with the heuristic approach. The parametrisation of the introduced regression functions can also directly express the subjective P_f–PC–FoS relationship directly, thereby sidestepping the inconsistencies that often accompany purely heuristic curve fitting.

We have also proposed a novel approach to estimating slope failure probability that focuses on the unknown unknowns to address our observations that previous dam failures were unexpected, and we expressed the subjective probability of failure as the probability that the pre-failure assessment was grossly incorrect in predicting stable conditions. We offer this path for other researchers to explore.

Quantified probabilities offer significant potential, particularly in areas where standards-based approaches are unavailable and for comparison and prioritisation purposes. However, owners making decisions over dam safety should recognise the inherent challenge of quantifying probabilities of one-off events, such as slope instability, magnified by incorporating outcomes from incomparable standards-based methods and by the inherently abstract nature of the subjectively estimated probabilities.

This is even more important in Australia where the legislative requirements are based on a different rationale, and the contemporary probability-based methods may neither be required nor sufficient to fulfill their legal obligations. As an alternative, we have proposed a simple, rational framework for addressing foreseeable risks associated with potential slope instability, inspired by the methods outlined in ICOLD Bulletin no. 194 (ICOLD, 2024).

9. ACKNOWLEDGEMENT

The authors would like to express their gratitude to Ryan Singh for his patience and insightful discussions throughout the development of this paper. We also extend our appreciation to Ryan Singh and James Thorp for their thoughtful review, valuable feedback, and inspiring comments.

CRediT authorship contribution statement

Jiri Herza: Conceptualization; Methodology; Visualisation; Writing – original draft; Writing – review & editing. **Hugo Fellows-Smith:** Formal analysis; Writing – original draft.

10. REFERENCES

Australian National Committee on Large Dams (ANCOLD) (2003). *Guidelines on Risk Assessment,* ANCOLD, Melbourne.

Australian National Committee on Large Dams (ANCOLD) (2022). *Guidelines on Risk Assessment*, ANCOLD, Melbourne.

Australian National Committee on Large Dams (ANCOLD) (2019). *Guidelines on Tailings Dams: Planning, Design, Construction, Operation and Closure*, ANCOLD, Melbourne.

Askari, H. & Mirakhor, A. (2015). Recurring financial crises—the causes, in *The next financial crisis and how to save capitalism, Palgrave Macmillan*, New York, pp. 16–34.

Baecher, G.B., Paté, M.-E. & de Neufville, R. (1980). Dam failure in benefit–cost analysis, *Journal of the Geotechnical Engineering Division*, vol. 106, no. 1, pp. 101–105.

Baecher, G. B., & Christian, J. T. (2003). *Reliability and statistics in geotechnical engineering*, John Wiley & Sons, Chichester.

Chang, M., Dalpatadu, R.J., Phanord, D. & Singh, A.K. (2018). A bootstrap approach for improving logistic regression performance in imbalanced data sets, *Matter: International Journal of Science and Technology*, vol. 4, no. 3, pp. 11–24.

Christian, J.T., Ladd, C.C. & Baecher, G.B. (1992). Reliability and probability in stability analysis, in *Stability and Performance of Slopes and Embankments II: Proceedings of the ASCE Specialty Conference*, vol. 2, American Society of Civil Engineers, New York, pp. 1071–1111.

Duncan, J. M. (2000). Factors of safety and reliability in geotechnical engineering. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(4), vol. 126, no. 4, pp. 307–316.

Euromoney 2015. SNB's shock move on the franc: lessons from Black Thursday, *Euromoney*, January, viewed 12 March 2025, https://www.euromoney.com/.

Government of Western Australia 2020, *Work Health and Safety Act 2020*, Government of Western Australia, Perth, viewed 12 March 2025,

Hartford, D. N. D., & Baecher, G. B. (2004). *Risk and uncertainty in dam safety*. Routledge, London.

Hicks, M., & Li, Y. (2018). Influence of length effect on embankment slope reliability in 3D. *International Journal for Numerical and Analytical Methods in Geomechanics*, 42(7), 891-915.

Herza, J., Coffey, J., & Singh, R. (2022). Key elements of tailings dam safety case. Proceedings of the 2022 ANCOLD Conference, Australian National Committee on Large Dams, Hobart.

Herza, J. & Singh, R. (2023). The risk of quantifying risk for tailings dams, in Management for Safe Dams*, Proceedings of the Symposium, 91st Annual Meeting of the International Commission on Large Dams*, Gothenburg, Sweden, 13–14 June 2023, International Commission on Large Dams, Paris.

Keelin, T. W. (2016). The metalog distributions, *Decision Analysis*, 13(4), 243–277, doi:10.1287/deca.2016.0338.

Lambe, T. W. (1985). Amuay landslides, Proc., *11th Int. Conf. on Soil Mechanics and Foundation Engineering*, Golden Jubilee Volume, San Francisco, Balkema, Boston, 137–158.

Laplace, P. S. (1774). Memoire sur les suites recurro-recurrentes et sur leurs usages dans la theorie des asards, *Memoire de l'Academie Royale des Sciences de Paris (savants etrangers)*, Tome VI, 353–371.

Mitchell, J. K., & Soga, K. (2005). Fundamentals of soil behavior (3rd ed.), Hoboken, NJ: Wiley.

Rana, N.M., Ghahramani, N., Evans, S.G., Small, A., Skermer, N., McDougall, S. & Take, W.A. (2022). Global magnitude–frequency statistics of the failures and impacts of large water-retention dams and mine tailings impoundments, *Earth-Science Reviews*, vol. 232, article 104144, doi:10.1016/j.earscirev.2022.104144.

Robertson, P. K., Rowe, R. K., & Olsen, K. B. (2019). Brumadinho dam failure: Lessons for tailings dam risk assessment and management, *Geotechnical News*, 37(3), 22–27.

Silva, F., Lambe, T. W., & Marr, W. A. (2008). Probability and risk of slope failure. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(12), 1691–1699.

Slivak v Lurgi (Australia) Pty Ltd [2001] HCA 6; 205 CLR 304, [53] (Gaudron J).

Smith, G. (2024, August 8). What is reasonably practicable – A work health and safety perspective [Webinar]. Engineers Australia.

The Economist (2015). The day the Swiss franc ceased to be boring, *The Economist*, 17 January, viewed 2 Jan 2025, https://www.economist.com/.

TÜV SÜD do Brasil (2018). Stability condition statement: Dam I, Córrego do Feijão Mine, *technical report prepared for Vale S.A.*, Belo Horizonte, Brazil, September 2018.

Vale S.A. (2019). Vale's statement on the Brumadinho dam collapse, *media release*, 29 January. Vale S.A., Rio de Janeiro, viewed 12 February 2025. https://vale.com/fi/w/clarifications-on-the-dam-i-of-the-corrego-do-feijao-mine.

Whitman, R. V. (1984). Evaluating calculated risk in geotechnical engineering. *Journal of Geotechnical Engineering,* 110(2), 143–188.

The Wall Street Journal (2014). Swiss National Bank faces pressure on franc cap, *The Wall Street Journal*, 5 December, viewed 10 December 2024, https://www.wsj.com/.

Xiao, Z., Lü, Q., Zheng, J., Liu, J., & Ji, J. (2020). Conditional probability-based system reliability analysis for geotechnical problems, *Computers and Geotechnics*, vol. 126, article 103751, doi:10.1016/j.compgeo.2020.103751