A practical framework for benchmarking failure rates of tailings storage facilities

H.L. Williams, A.B. Leik & D.E. Hockley

SRK Consulting, Vancouver, Canada

ABSTRACT: The probability of failure for a tailings storage facility (TSF) depends on dam characteristics, site conditions, and the quality of design, construction, operations, and governance. Only some of these factors, most notably construction method and operational status, are systematically captured in global inventories and failure databases. This paper provides a practical statistical framework to benchmark annual TSF failure rates by construction method (e.g. upstream, downstream, single-raise, dry-stack) and operational status (active, inactive, closed).

The framework combines global compilations of TSF failures with a large inventory of non-failed facilities and explicitly accounts for dam years of exposure. A hierarchical Bayesian structure is used to "share" information across related categories, so that failure rates can be estimated even for facility types with very few or no recorded failures, while still reflecting the uncertainty.

The result is a set of annual failure rate estimates with 95% credible intervals to show the uncertainty. The analysis confirms several robust patterns: active facilities fail more often than inactive and closed facilities; upstream TSFs have higher estimated failure rates than downstream facilities; and single-raise and dry-stack TSFs show lower estimated rates, but with wider uncertainty bands due to limited failure history. Sensitivity tests indicate that these overall rankings are stable for data-rich categories, but that estimates for data-sparse categories such as dry-stacks are sensitive to how many failures are ultimately recognized.

For practitioners, the framework provides benchmark failure rates that can be used directly in screening-level and semi-quantitative risk assessments. Because the approach is updateable as new data becomes available and can link with physically based causal models, such as Bayesian networks, to represent how governance, design quality, and site conditions modify baseline risks at individual facilities, it provides a foundation for progressively improving risk-informed TSF management.

1 INTRODUCTION

The probability of failure for a tailings storage facility (TSF) depends on many factors including dam characteristics (e.g. construction method, height), site conditions (e.g., climate, seismicity, foundation geology), and the quality of design, construction, operations, and governance. While some factors, like construction method and operational status, are documented in publicly available dam inventory and failure databases, others, such as governance quality, are not readily accessible in existing datasets. This limitation means we need to start with baseline failure rate estimates based on available characteristics which can then be refined to incorporate other information sources and engineering judgement.

This paper presents a hierarchical Bayesian framework for estimating baseline TSF failure rates by construction method and operational status, and quantifying the associated uncertainty. The remaining sections describe the data sources and methodology (Section 2), present baseline failure rate estimates with credible intervals (Section 3), compare the hierarchical approach to simpler alternatives and assess robustness (Section 4), and conclude with recommendations for practitioners and future research (Section 5).

2 METHODOLOGY

2.1 *Data sources and processing*

Tailings dam information was compiled from the following sources:

- Failures: The ICOLD (2019), Rana (Rana et al, 2022), World Mine Tailings Failures (WMTF, 2025) and Wise (2025) tailings failure databases were combined with three additional failures at Morowali Industrial Park (IMIP) in Indonesia in 2025 to 2026. The dataset has a total of 181 failures with a known construction method, excluding those classified as "incidents" (in other words, those that did not feature significant tailings release or loss of structural integrity).
- Non-Failures: Franks (Franks et al, 2021) compiles tailings disclosure data for 1743 non-failed TSFs and includes information on construction method, operational status, facility age, and location. Despite not being a complete inventory, the sample includes geographic regions and regulatory environments representative of the global TSF population.

Data completeness varies. Construction method is documented for 77% of TSFs (31% failures and 88% non-failures) and operational status is documented for 91% (69% failures and 96% non-failures).

The following processing assumptions were applied for data consistency:

- Removed duplicate failure records using a two-stage process: first, automated matching identified obvious duplicates (same failure date, location, and similar facility name or ID); second, remaining potential duplicates with similar dates and locations but different names (e.g., Brumadinho/Feijão, Samarco/Fundão) were reviewed manually.
- Reported dam types and operational statuses were recoded into standardized categories using the case-insensitive groupings shown in Tables 1 and 2.In this paper, "active" facilities are those currently receiving tailings, undergoing construction raises, or being re-mined; "inactive" facilities have ceased deposition but are not yet fully stabilized or reclaimed; and "closed" facilities have completed deposition and are reported as closed or reclaimed.
- Centreline facilities were not analysed as a separate category because they are rarely reported explicitly and the definition of "centreline" varies between sources; facilities described with any upstream component were conservatively classified as upstream.
- TSFs with unknown construction method or operational status were excluded from category-specific summary statistics.
- Two WMTF records coded as inactive were reclassified based on independent research. Merriespruit No. 4A (South Africa, 1994) and Marcopper Tapian Pit (Philippines, 1996) were recoded as Active at the time of failure.

Table 1 – Mapping of reported construction method descriptions to standardized dam type categories.

Assigned Dam Type	Dam Types in Database
Upstream	Upstream, US, US paddock, DS then US, DS / US
Downstream	Downstream, DS
Single Raise	Single raise, NA, ring, ring-dyke, unraised
Dry-stack	Drystack, Dry-stack, filtered

Table 2 – Mapping of reported operational status descriptions to standardized status categories.

Assigned Status	Dam Statuses in Database
Active	Active, re-mining, construction
Inactive	Inactive, inactive care and maintenance, inactive/reclamation under planning, in partial reclamation, care and maintenance, ongoing reclamation, inactive / opened for harvesting
Closed	Closed, reclaimed, rehabilitated

2.2 *Key assumptions*

Key assumptions:

- Average TSF failure frequency: We assume an average of 4.3 TSF failures per year globally, consistent with the long-term average reported in Rana (2022). Combining this with the global TSF population estimates in the next bullet establishes a target baseline failure rate of approximately 2.1E-4 per dam-year. (Rana's full analysis shows that failure frequencies have varied over time. We adopt a single value for simplicity and because our primary interest is in comparing relative rates of failure rather than estimating absolute rates.)
- Global TSF population estimates: The base case uses Rana's (2022) estimate of 20,230 TSFs. We also test sensitivity using a range of published estimates from 6,251 (Stark et al 2022) to 32,000 (high estimate from Rana 2022). While specific failure rate estimates scale with population assumptions, the relative differences between construction methods and operational status are unchanged.
- TSF years extrapolation: estimating failure rates requires an assumption of exposure time, measured in dam-years, for each construction method and status category. One dam-year represents one facility operating in a given status category for one year. Because the non-failure database is incomplete, we extrapolate total exposure (T_i) for each category (i) from the observed exposure (T_i^{obs}) and estimated total dam years (T_T) from the relative exposure patterns in the sample database using Equation 1.

$$T_i = \frac{T_i^{obs}}{\sum_i T_i^{obs}} \times T_T \quad (1)$$

Where T_i^{obs} is the observed dam-years for a given category i, $\sum_i T_i^{obs}$ is the total observed dam years across all categories, and T_T is the estimated total dam years experience taken as the total dam years experience summed from Rana's (2022) median TSF construction histograms. We estimate T_T ~ 450,000 dam-years for Rana's base case estimate of 20,320 TSFs but note that the T_T estimate scales with different total population assumptions.

- Redistribution of dam-years by operational status: the databases report construction year and current operational status, but not status history. We first calculate each facility's age as the inventory year minus the construction year. For active facilities, this age is assigned entirely to the active category. For inactive facilities, we assume an initial active period equal to the mean active age for that construction type and allocate the remaining years as inactive. For closed facilities, we assume typical active and inactive durations before closure and assign any remaining years to the closed category. This avoids over-allocating exposure to inactive and closed facilities based on only their current status.

Calculations and analysis framework

This analysis adopts survival analysis methods for censored data, recognizing that a TSF with no recorded failure may still fail in the future. The hierarchical Bayesian model represents failure counts with Poisson likelihoods conditional on exposure time. Readers are directed to NASA (2011) for a more detailed description of this methodology.

The key advantage of the hierarchical structure is information sharing through "partial pooling": categories with sparse data borrow strength from related categories through shared parameters, while data-rich categories are driven mainly by their own observations. This allows us to learn construction-method and status effects from the full dataset and propagate them to categories with few or no observed failures (e.g., dry-stack facilities).

The model uses Gamma-Poisson conjugate priors, a standard choice in reliability engineering (NASA, 2011) that allows closed-form solutions that can be solved directly without the need for complex numerical simulation.

The hierarchical model has three levels, as shown in Figure 1, which also provides an overview of the calculation process described in the following subsections.

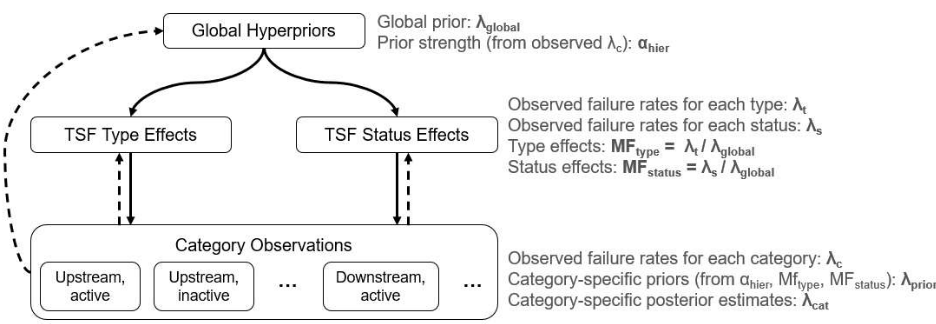


Figure 1 – Overview of hierarchical model framework with arrows indicating direction of information sharing.

Step 1: compiling observations

Observations include observed failures (k_c) and exposure in dam-years extrapolated to match global TSF estimates (T_c) for each category (unique construction method and status combination). Observations are also combined ("pooled") by construction method and status to estimate multiplication factors later.

Step 2: global hyperparameters

Our global prior average failure rate, λ_{global}, is calculated using Equation 2.

$$\lambda_{global} = 4.3 \,/\, N_{TSF} \qquad (2)$$

Where 4.3 is the historical number of TSF failures per year and N_{TSF} is the assumed global population of TSFs.

Having specified this global average, we then need to decide how strongly it should influence the estimates for each individual category. This "weighting" of the global prior is controlled by the hyperparameter, α_{hier}. We estimate α_{hier} from the observed variation in failure rates across all categories using the method of moments (Equation 3). Large variations in observed rates between categories leads to a smaller α_{hier} (typically less than 1), indicating that a "typical" rate (λ_{global}) does not represent the data well and the model should rely more on the category-specific observations. Smaller variation leads to a larger α_{hier}, indicating that λ_{global} is a reasonable summary and should carry more weight.

$$\alpha_{hier} = \frac{\mu^2}{s^2} \quad (3)$$

Where μ and s^2 represent the mean and variance across all observations

Step 3: estimating type (construction method) and operational status effects

Next, multiplication factors (MF_i) are estimated for each type of construction method and operational status using Equation 4:

$$MF_i = \frac{\lambda_i}{\lambda_{global}} = \frac{\frac{k_i}{T_i}}{\lambda_{global}} \qquad (4)$$

Where k_i is the number of observed failures for each construction method and status, and T_i, are extrapolated exposure in dam-years for each construction method and status.

Step 4: category-specific estimates

The category-specific prior mean ($\lambda_{prior,c}$) is calculated using Equation 5 for each combination of construction method and status.

$$\lambda_{prior,c} = \lambda_{global} \times MF_{type} \times MF_{status} \quad (5)$$

Which is then used to estimate a posterior distribution of the failure rate (Equation 6), allowing any specific probability to be calculated using the gamma inverse distribution function.

$$\lambda_c \sim Gamma\left(\alpha_{hier} + k_c,\ \beta_c + T_c\right) \quad (6)$$

where $\beta_c = \alpha_{hier}/\lambda_{prior,c}$, and mean failure rates can be calculated from Equation 7.

$$\mu_{\lambda_c} = \frac{\left(\alpha_{hier} + k_c\right)}{(\beta_c} + T_c) \qquad (7)$$

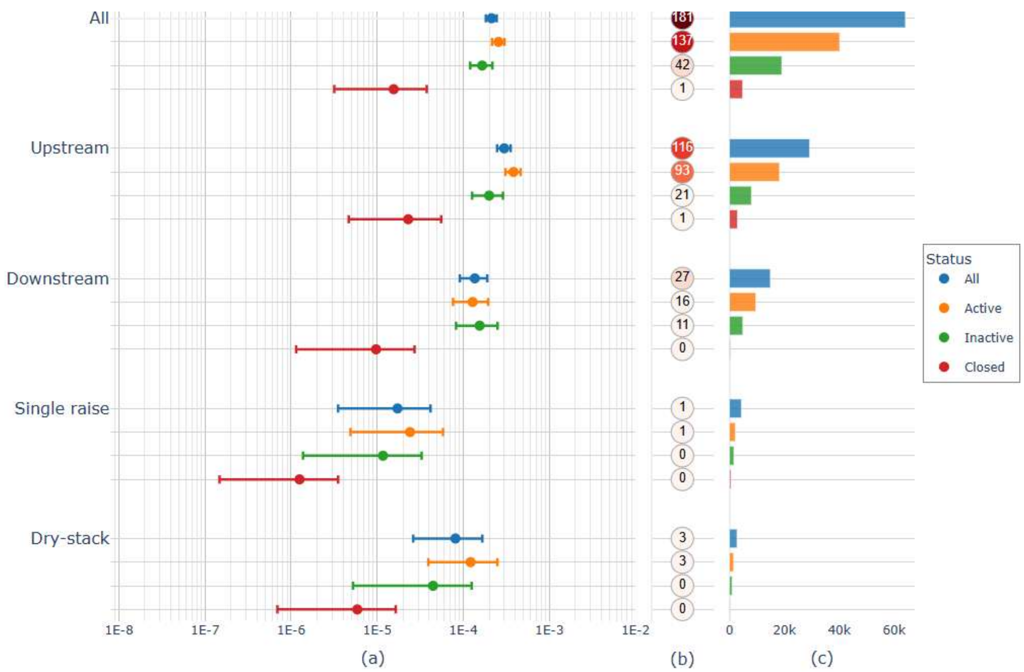
Because each category-specific failure rate λ_c has a Gamma posterior, 95% credible intervals are obtained directly from the 2.5th and 97.5th percentiles of that Gamma distribution:

$$CI_{95\%} = \left[\Gamma^{-1}\left(0.025,\ \alpha_{hier} + k_c,\ \beta_c + T_c\right), \Gamma^{-1}\left(0.975,\ \alpha_{hier} + k_c,\ \beta_c + T_c\right)\right]$$
(8)

Where $\Gamma^{-1}(p;)$ denotes the pth quantile of the Gamma distribution with the indicated shape and rate parameters.

3 RESULTS

Figure 2 shows the estimated annual TSF failure probabilities by construction method and operational status with 95% credible intervals (plot a). The remaining two plots show the observed failure counts (b) and documented dam-years of exposure (c), illustrating data availability. While specific failure rates scale with global TSF population assumptions, the overall patterns of relative differences between categories and widths of uncertainty bounds remain constant for different assumptions.



.

Figure 2 – Results by construction method and status for a base case assumption of 20,230 TSFs: (a) Estimated TSF failure rates with 95% credible intervals, (b) observed number of failures, and (c) observed dam-years

Table 3 presents the results as relative failure-rate multipliers which are independent of population assumptions. It shows how each construction method and operational status modifies the baseline failure rate, where the baseline rate could range from 1.3E-4 to 6.9E-4 depending on TSF population assumptions. The 95% credible intervals are shown in brackets representing the uncertainty in each estimate.

Table 3 – Relative change to baseline failure rate assumption by TSF construction method and status

Dam Type	All	Active	Inactive	Closed
All	1.0 *(0.86 to 1.2)*	1.2 *(1.0 to 1.4)*	0.78 *(0.57 to 1.0)*	0.07 *(0.01 to 0.18)*
Upstream	1.4 *(1.2 to 1.7)*	1.8 *(1.46 to 2.18)*	0.94 *(0.60 to 1.4)*	0.11 *(0.02 to 0.26)*
Downstream	0.64 *(0.43 to 0.89)*	0.60 *(0.36 to 0.91)*	0.73 *(0.39 to 1.2)*	0.05 *(0.01 to 0.13)*
Single Raise	0.08 *(0.02 to 0.20)*	0.11 *(0.02 to 0.27)*	0.05 *(0.01 to 0.15)*	0.01 *(0.001 to 0.02)*
Dry-stack	0.38 *(0.12 to 0.78)*	0.57 *(0.18 to 1.2)*	0.21 (0.02 to 0.59)	0.03 *(0.003 to 0.08)*

Note: Results are multiplicative factors relative to the baseline annual failure rate. For example, 2.0 means twice the baseline failure rate and 0.05 means 5% of the baseline failure rate.

Three main patterns emerge from these results:

- **Operational status effects:** Generally, active facilities consistently have the highest failure rates, and closed facilities the lowest. The status effect multipliers are approximately 1.2 for active facilities (~20% higher than the portfolio average); 0.78 for inactive facilities (~22% lower), and 0.07 for closed facilities (~93% lower). Within the downstream category, however, the estimated failure rate for inactive facilities is slightly higher than for active facilities. This is discussed further in Section 4.4.
- **Construction method differences**: Upstream construction shows the highest estimated failure rates across all operational statuses, with estimated failure rates for active upstream facilities being approximately 1.8 times higher than the base rate. Downstream and single raise facilities have lower estimated rates, typically between one-half and two-thirds of the base rate. Dry-stacks also show lower estimated failure rates, however, these estimates are based on limited data and have wide credible intervals.

- **Uncertainty quantification**: The widths of the credible intervals vary significantly, reflecting the differences in data availability and experience with specific facilities. Data-rich categories with many observed failures and dam-years of experience, such as active upstream TSFs, have relatively narrow intervals. Dry-stack facilities, which have fewer dam-years and only three recorded failures, have credible intervals that span about an order of magnitude, reflecting the greater uncertainty in their long-term performance.

4 DISCUSSION

4.1 *Comparison to alternative approaches*

To demonstrate the advantages of the hierarchical Bayesian approach, we compare it to three alternative approaches below that progressively increase in sophistication.

Example 1: Percentage-based estimate (ignoring exposure time)

A common approach is to estimate failure probabilities from observed percentages, ignoring exposure time, as in Equation 8:

$$P(Category) = \frac{P(Category\,|Failure) \times P(Failure)}{P(Category)} \qquad (8)$$

For single-raise facilities, this produces reasonable values for the data-rich active category, but is incomputable for inactive and closed facilities because it would require dividing by zero. More fundamentally, it ignores exposure time and uncertainty.

Example 2: Simple pseudo-counts

A second approach uses pseudo-counts, representing a belief that "we should assume at least one potential failure" to avoid zero counts. Table 4 shows the resulting failure rate estimates when an arbitrary pseudo-count of 1 is added to the inactive and closed single raise categories with zero failure counts. We compare two pseudo-count approaches: percentage-based (Method 2A), and a simple rate-based approach that divides failures by extrapolated dam years (Method 2B).

As shown in Table 4, this method produces unrealistically high estimates for facilities with sparse data. For closed facilities, the pseudo-count approaches (Methods 2A and 2B) produce failure rate estimates similar or higher to base rates despite zero recorded failures. The choice of pseudo-count (0.5? 1? 2?) is also arbitrary and can dramatically affect the results.

Example 3: Independent (non-hierarchical) Bayesian model

A key difference between independent and hierarchical Bayesian models is that independent models apply priors separately to each category. This means that each estimate depends only on that category's own observed failures and exposure with no consideration of trends from related categories.

Continuing to use single raise facilities as an example, an independent model would estimate higher rates for closed facilities than inactive facilities because closed single-raise facilities have zero failures but less documented exposure (~7,500 extrapolated dam-years versus ~22,000 for inactive). The hierarchical model incorporates the strong evidence from all construction methods that closed facilities fail substantially less frequently, producing lower estimates for closed single raise facilities. The uncertainty associated with the reduced exposure time is captured through wider credible intervals.

Table 4 – Comparison of failure rate estimation methods for single-raise facilities (pseudo-count = 1)

Method	Active Single-Raise	Inactive Single Raise	Closed Single Raise
Recorded failures	1	0	0
Recorded dams	58	44	26
Extrapolated dams	*701*	*532*	*314*
Recorded dam-years	2,172	1,645	561
Extrapolated dam-years	*~29,000*	*~22,000*	*~7,500*

1) Percentage-based Bayes (no pseudo-count)	2E-5	**N/A**	**N/A**
2A) Pseudo-count=1 with percentage-based Bayes	2E-5	**6E-5**	**3E-3**
2B) Pseudo-count=1 with rate-based estimate	9E-4	6E-4	**2E-3**
3) Independent Bayesian model	5E-5	2E-5	**5E-5**
4) Preferred Bayesian hierarchical method	2E-5	1E-05	1E-6

4.2 *Robustness of estimates*

Categories with zero or very few documented failures naturally raise the question of how much a single additional failure would change the estimated rates. This was especially relevant for dry-stack facilities. When the original analysis was first completed, no dry-stack structural failures were recorded in any of the compiled databases, reinforcing the common view that dry-stacks are the least failure-prone construction method. Since then, three dry-stack failures have been confirmed, all at Morowali Industrial Park (IMIP) in Indonesia. These events have prompted debate about whether they are indicative of unreported dry-stack failures elsewhere, a local anomaly not representative of global performance, or evidence that dry-stacks are roughly as likely to fail as conventional TSFs.

Figure 3 shows how the estimated dry-stack failure rate varies under different assumed failure counts, compared with upstream facilities. The blue points show upstream estimates, which are well constrained by many events and changes very little even when several failures are added or removed. The red points show dry-stack estimates, which are much more sensitive to failure counts.

The top point corresponds to the zero-failure case prior to the IMIP events; its wide credible interval lies entirely below the downstream and portfolio averages. Treating the IMIP cluster as one or two independent failures shifts the estimate right so that the upper bound of the credible interval just touches the downstream mean, meaning a "dry-stacks are definitively safer" conclusion is no longer statistically robust. Counting all three IMIP failures as independent (current base case) moves the mean further right and the 95% interval now overlaps both downstream and portfolio averages. The remaining points show how the estimate would change if additional, as-yet unreported dry-stack failures were recognized. Adding roughly four to eight failures would move the dry-stack mean into the same range as downstream, with substantial overlap in credible intervals, and only when the total reaches the low double digits would the mean approach current upstream rates.

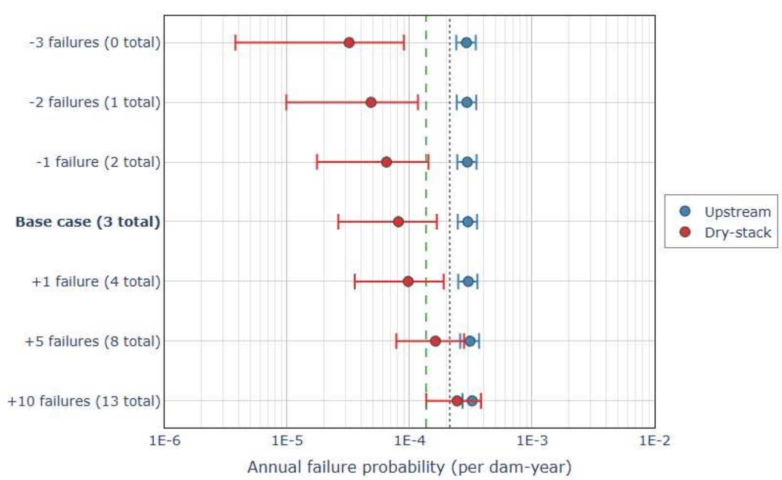


Figure 3 – Sensitivity of failure rate estimates to additional failures for upstream (blue) and dry-stack facilities (red). Two dashed reference lines represent average failure rate estimates for all facilities (grey) and downstream facilities (green)

Two conclusions can be drawn from this. First, upstream remains clearly the highest-failure-rate construction method and its estimated rate is insensitive to the exact failure count. Second, the ranking of dry-stack relative to downstream and the portfolio average is very sensitive to the number of failures. With zero or one failures, dry-stacks appear clearly better; with two or three failures the ordering becomes uncertain; and with around eight or more failures the most likely dry-stack rate would be similar to, or potentially higher than, the downstream average. The current evidence therefore does not support strong claims that dry-stacks fail less often than conventional slurry impoundments but also does not yet prove the conclusion that they fail more often on average.

Based on a similar sensitivity analysis performed for all categories, we can draw the following conclusions:

- Stability of data-rich categories: The total (all facility) baseline estimates are robust for all operational statuses, as are the baseline estimates for active and inactive upstream and downstream facilities
- Construction method hierarchy: Upstream construction shows higher failure rates than downstream construction. This would remain true even if about twenty failures were missing from the downstream category.
- Operational status hierarchy: On average across all construction methods, active facilities show higher rates than inactive or closed facilities. This would remain true unless inactive/closed failures are systematically underreported by about an order of magnitude. The main exception is downstream TSFs, where inactive facilities have a slightly higher estimated rate than active facilities. A possible explanation is discussed in Section 4.4.
- Data-sparse category interpretation: Dry-stack and single-raise methods show low observed failure rates, but also have more limited operational history or might be less consistently documented in the available databases. With three dry-stack failures (all at a single industrial park) and only four single-raise failures, the estimates have wide uncertainty intervals. A conservative interpretation is that the current data show no demonstrated elevation in failure rate for dry-stacks relative to conventional methods, but also do not yet demonstrate a lower failure rate. As operational experience accumulates and reporting improves, especially for increasingly common dry-stack facilities, future updates will clarify whether the estimated rates represent genuine performance differences or data limitations.

4.3 *Failure probability vs risk*

It is important to distinguish between probability of failure and risk. The benchmarks in this paper quantify failure rates, but risk also depends on the consequences of failure. Two facility types can have similar failure rates but very different risks if the expected consequences differ. Recent dry-stack failures provide a useful example: they highlight uncertainty about long-term performance and emphasize that strong design, construction, and operational controls remain just as essential as for other tailings construction methods; however, filtered tailings generally have lower mobility and runout potential than slurry tailings, so the consequences of failure are likely to be less severe.

4.4 *Physical mechanisms explaining the differences*

Interpretation of differences

The observed differences in failure rates across construction methods and operational statuses (see Figure 2 and Table 3) reflect underlying physical mechanisms and risk factors. While a comprehensive causal analysis is beyond the scope of this paper, we can identify several plausible explanations supported by engineering principles and failure case histories.

- Status mechanisms: Active facilities experience continuous loading from construction and tailings deposition, evolving geometry and stress conditions, and increased operational issues such as poor water management. Inactive facilities eliminate many common operational issues but may not yet be fully stabilized for the long-term. Closed facilities benefit from stable geometries, long-term water management systems, and protective covers. Downstream facilities provide an interesting nuance: their estimated failure rate is higher for inactive than for active facilities. One hypothesis is that downstream embankments are relatively resilient to short-term operational variability during active deposition, but become more vulnerable once deposition ceases and attention and resources shift elsewhere. Several inactive downstream failures in the database involve deferred maintenance, deteriorating water management, or delayed implementation of closure works, consistent with this interpretation.

- Construction mechanisms: Elevated upstream risks can be attributed to several factors, including higher liquefaction potential from raising on previously deposited tailings, typically higher phreatic surfaces, and more challenging operational requirements. Downstream construction generally provides more resilience to operational issues and reduced liquefaction susceptibility. Single-raise eliminates many construction risks, and dry-stack facilities, which dewater tailings to specified moisture content ranges before placement, theoretically eliminate liquefaction mechanisms when properly implemented.
- Failure mechanisms: A review of the failure database indicates that many failures can be attributed to some combination of misoperation (e.g. poor water management), unidentified layer (e.g. foundation issues), damage (e.g. blockage of a critical water management structure, streamflow erosion of a dam toe), construction issues (e.g. high rate of rise, inadequate borrow material, steeper than design), and / or inappropriate design, with some mechanisms being more concentrated in specific categories of facilities. For example, foundation issues were concentrated in upstream facility failures, construction issues in active facility failures, and misoperation in active upstream facility failures. Many failures involve multiple factors with inadequate governance as a common amplifier.

This shows that the observed statistical patterns related to construction-method and status effects can be grounded in physically based failure mechanisms, increasing confidence that the patterns reflect genuine risk factors rather than statistical artifacts.

4.5 *Future Extension: Bayesian network for causal modelling and risk assessment*

Many readers may note that the hierarchical Bayesian framework developed in this paper is well suited to implementation within a Bayesian network. Such a network could formalize the statistical findings presented here within a broader causal structure that explicitly represents additional factors such as governance quality, design adequacy, site conditions, and other dam-specific characteristics.

Embedding the benchmark failure rates in a Bayesian network would allow the analysis to move beyond generic, portfolio level estimates toward project specific risk assessments that account for combinations of factors that can make an individual significantly more or less risky than the average. The development and application of such a causal Bayesian network is underway and will be presented in a second paper.

5 CONCLUSIONS AND RECOMMENDATIONS

This paper presents a hierarchical Bayesian framework for benchmarking annual failure rates of tailings storage facilities by construction method and operational status. By combining global failure compilations with non-failure inventories and explicitly accounting for exposure time, the framework generates baseline failure rate estimates with credible intervals that reflect data limitations and uncertainty. The hierarchical structure provides defensible estimates even for newer construction methods, such as dry-stacks, with few or no documented failures.

The results confirm several robust portfolio-level patterns:

- On average, active facilities exhibit higher failure frequencies than inactive and closed facilities, although downstream TSFs show slightly higher rates in the inactive phase.
- Upstream construction shows higher estimated failure rates compared to downstream.
- Single-raise and dry-stacks show lower estimated failure rates compared to other construction methods. However, the data is sparse and credible intervals are wide. For dry-stack facilities, three failures, all at one Indonesian industrial park in 2025 to 2026, now provide an initial empirical estimate, but they are too few and too geographically concentrated to support strong inferences about global performance. A conservative interpretation of these results is that the evidence supports "no demonstrated elevated failure rate" rather than "demonstrated lower failure rate".
- Sensitivity analyses indicate that the overall ranking of construction methods and operational statuses is stable under plausible numbers of additional failures or missing

events for data-rich categories such as upstream and downstream facilities, but that estimates for data-sparse categories such as dry-stacks remain sensitive to how many failures are ultimately recognized.

Beyond providing benchmark failure rates, the framework is designed to integrate with physically-based causal models, such as Bayesian networks, that represent how governance, construction quality, operation, and site conditions modify baseline failure probabilities and risks at individual facilities. In practice, practitioners can use the benchmark rates as starting points for screening-level and semi-quantitative risk assessments and then update them with site-specific indicators where information is available.

The Bayesian methodology allows straightforward updating as new information becomes available or assumptions evolve, providing a "living" model that can improve with industry experience. Future work should focus on refining and expanding the set of explanatory variables (e.g. specific dam characteristics, level of governance, climate, seismicity) and their linkages to develop improved tools that translate portfolio-level statistics into facility-specific risk estimates suitable for decision-making and risk prioritizations.

6 REFERENCES

Franks, D.M., Stringer, M., Torres-Cruz, L.A. et al. 2021. Tailings facility disclosures reveal stability risks. Sci Rep 11, 5353. https://doi.org/10.1038/s41598-021-84897-0

ICOLD (International Commission on Large Dams). 2019. Bulletin 188. Incident Database—Bulletin 99 Update: Statistical Analysis of Dam Failures (Final Draft, December). Paris, France.

ICOLD (International Commission on Large Dams). 2025. World Register of Dams. https://www.icold-cigb.org/GB/world_register/general_synthesis.asp (accessed November 2025).

NASA. 2011. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Second Edition. NASA/SP-2011-3421.

Rana, N.M., Ghahramani, N., Evans, S.G. et al. 2022. Global magnitude-frequency statistics of the failures and impacts of large water-retention dams and mine tailings impoundments. Earth-Science Reviews 232, 104144. https://doi.org/10.1016/j.earscirev.2022.104144

Stark, T. D., Moya, L. & Lin, J. 2022 Rates and Causes of Tailings Dam Failures, Advances in Civil Engineering, 2022, 7895880, 21 pages. https://doi.org/10.1155/2022/7895880

WISE, 2025. Chronology of Major Tailings Dam Failures. https://www.wise-uranium.org/mdaf.html (accessed November 2025).

World Mine Tailings Failures (WMTF). 2025. Database. Downloaded 15 April 2025. https://worldminetailingsfailures.org/