

Beyond volume proxies: An exposure-based scenario-screening framework for conditional potential loss of life in tailings storage facility portfolios

Vancho Adjiski^{a,*}, Tomasz Orpiszewski^b, Peter Schwendner^b

^a Faculty of Natural and Technical Sciences, Goce Delcev University, Stip, North Macedonia

^b Zürich University of Applied Sciences, School of Management and Law, Winterthur, Switzerland

ARTICLE INFO

Keywords:

Tailings storage facilities
Tailings dam failure
Conditional consequence screening
Potential loss of life
Downstream population exposure
Scenario-set modelling

ABSTRACT

Tailings storage facilities (TSFs) are expanding, and occasional failures can cause severe downstream impacts. This study presents a screening-level framework for conditional Potential Loss of Life (PLL, given failure) using open datasets, Google Earth Engine (GEE), and an Excel release model. The workflow estimates dam geometry by DEM cut-and-fill, caps release at 0.90 of storage, routes a MERIT Hydro-guided corridor, applies exponential thickness decay with minimum-thickness trimming, and evaluates PLL on the WorldPop 100 m grid.

Each site is evaluated using 144 equally weighted deterministic combinations and summarised with scenario-exceedance curves, scenario-set Value-at-Risk at 95% (VaR95) and scenario-set Expected Shortfall at 95% (ES95). Across 17 sites, upper-tail PLL spans nearly three orders of magnitude, with VaR95 from 0.1 to 148 and P99 from 0.15 to nearly 200. A high-consequence group emerges, while most remain low. Scenario-exceedance curves reveal varied distributions, including long-tail behaviour where typical outcomes are low but some combinations produce severe PLL. Partial rank analysis showed a stronger independent association of PLL with exposed population ($\rho_{\text{partial}} = 0.64$) than with stored volume ($\rho_{\text{partial}} = 0.16$), although the 17-site sample limits generalisation.

The framework supports standardised ranking and prioritisation, complementing site-specific breach hydraulics and evacuation analyses. Results should be interpreted within assumptions on routing, release estimation, transferable depth-fatality curves, population uncertainty and equal design weighting. Estimates represent conditional consequences, not probabilistic annual risk. Future work should extend beyond PLL, which captures only fatality consequences and cannot substitute for assessing displacement, livelihoods, environmental harm or wider social consequences.

1. Introduction

Mine tailings, the fine-grained waste from ore processing, are generated in increasing quantities as demand grows for metals used in low-carbon energy systems. Global tailings generation is approximately 13 billion tonnes annually, while meeting 2020–2050 demand for key energy-transition commodities could generate up to 2 trillion tonnes of tailings and waste rock combined [1]. This growth

* Corresponding author.

E-mail address: vanco.adjiski@ugd.edu.mk (V. Adjiski).

expands inventories of tailings storage facilities (TSFs) and requires transparent portfolio-level risk governance [1–3]. TSF failures remain infrequent but high-consequence events [4,5]. Since 1915, 257 failures have been identified, with more than half of total released volume occurring after 2000 [1,6]. Brumadinho in 2019 and Mount Polley in 2014 illustrate consequences ranging from rapid downstream propagation and fatalities to infrastructure damage and persistent contamination [7–9].

These events have accelerated governance and disclosure initiatives, including the Global Industry Standard on Tailings Management (GISTM) and the Global Tailings Portal [2,3,10,11]. GISTM requires credible failure scenarios and assessment of downstream exposure, vulnerability, and consequences across loss of life, environmental, health, social, cultural, infrastructure, and economic dimensions [2]. However, Hopkins and Kemp [12] show that credibility also depends on industry power, accountability, independent assurance, and whether external actors can interrogate operator-generated assessments. Existing classification schemes are not always directly comparable or accompanied by a common quantitative PLL calculation [13–15]. Facilities within the same category can therefore have different downstream exposure and fatality potential [16,17]. Portfolio stakeholders need a rapidly applicable method that provides comparable, quantitative, and uncertainty-aware conditional PLL estimates across facilities [3,10,15,18].

Simple facility-size proxies are also insufficient [4,5,19]. Relationships between stored and released tailings volumes are widely used for preliminary release estimation, but their predictive strength decreases with larger failure datasets. Piciullo et al. [6] reported that the coefficient of determination for released versus stored volume decreased to $r^2 \approx 0.59$ in a larger sample, compared with values close to 0.9 in earlier studies, and that the average released-to-stored volume ratio was approximately 0.27. Consequences therefore depend not only on release magnitude but also on downstream routing and population exposure [15,17]. Conversely, detailed breach-hydrodynamic models require site-specific breach, rheological and geotechnical inputs and are often impractical for initial screening of large portfolios [11]. A phased approach is therefore needed in which standardised screening identifies facilities requiring detailed modelling and field verification [10,15].

The specific methodological gap is the absence of a scalable, reproducible open-data workflow combining regression-based release scenarios, downstream routing, spatially explicit population exposure, depth-dependent fatality estimation and structured sensitivity analysis to generate comparable conditional PLL metrics across TSF portfolios. Existing approaches cover only parts of this sequence, provide categorical outputs, or require detailed site-specific inputs. Table 1 positions the proposed framework relative to representative alternatives. Its contribution is integrative rather than a new breach-hydraulics solver or fatality function: it links release estimation, hydrography-guided routing, population exposure, conditional PLL and upper-tail diagnostics in one portfolio-scale workflow.

To address this gap, we developed a Google Earth Engine (GEE) framework coupled with an Excel-based release-volume model. Global elevation, hydrography and population datasets are combined with simplified, parameterised flow-thickness routing to estimate downstream exposure and depth-dependent PLL. Release and routing uncertainty is explored through a full-factorial set of 144 deterministic parameter combinations per site, interpreted using equal design weights and summarised with scenario-exceedance curves, scenario-set Value-at-Risk at 95% (VaR95), and scenario-set Expected Shortfall at 95% (ES95).

Applied to 17 TSFs, the framework revealed substantial variation in upper-tail conditional PLL and a stronger independent rank association with downstream population exposure than with estimated stored volume. It is intended to prioritise subsequent investigation, not to estimate annual failure risk or replace site-specific breach-hydraulic, evacuation or emergency-response studies.

2. Methodology

We present an end-to-end workflow for screening the consequences of tailings dam failure, implemented in GEE and coupled to a lightweight Excel regression (Fig. 1). Throughout the workflow, H denotes dam height, V_{dam} the estimated stored volume, F the cap-aware release-fraction set, $f \in F$ a selected released fraction, V_r released volume, L the e-folding length, t_{min} the minimum meaningful thickness and $halfW$ the corridor half-width. All mathematical symbols are defined immediately after the equation, or closely related group of equations, in which they first appear, together with their physical meaning and units. The implementation uses two companion GEE scripts that follow the same sequence of operations: an interactive script (model-1-GEE) for site setup and diagnostics, and a batch script (model-2-GEE) for evaluation of the full scenario set and statistical aggregation. In model-1-GEE, the dam polygon is used to compute dam height H and stored volume V_{dam} by digital elevation model (DEM) cut-and-fill [21,22]. These scalars (V_{dam} , H) are passed to the Excel model, which returns a cap-aware release set $F = \{f_{16}, f_{50}, f_{mid}^{dyn}, f_{hi}^{dyn}\}$ from a log-space regression with residual

Table 1
Comparison of representative approaches for TSF consequence assessment.

Approach and references	Main output	Limitation for portfolio-scale PLL screening
Empirical release and runout models [6,19,20]	Released volume, release fraction or runout estimates derived from historical failures	Do not translate release estimates into spatially explicit population exposure and PLL
Disclosure-based portfolio assessments [10]	Facility characteristics and reported consequence or hazard classes	Outputs are categorical, scheme-dependent and not based on a harmonised PLL calculation
Empirical loss-of-life methods [16,18]	Population at risk and potential fatalities for an established inundation area	Require an independently derived inundation footprint and were developed mainly for water-dam applications
Detailed site-specific modelling [7]	Breach hydraulics, non-Newtonian runout, exposure, warning and evacuation outcomes	Requires extensive site-specific data and substantial modelling effort
Proposed Google Earth Engine–Excel framework	Routed exposure, conditional PLL distributions and upper-tail scenario diagnostics	Uses simplified corridor and thickness representations and does not replace detailed hydraulic modelling

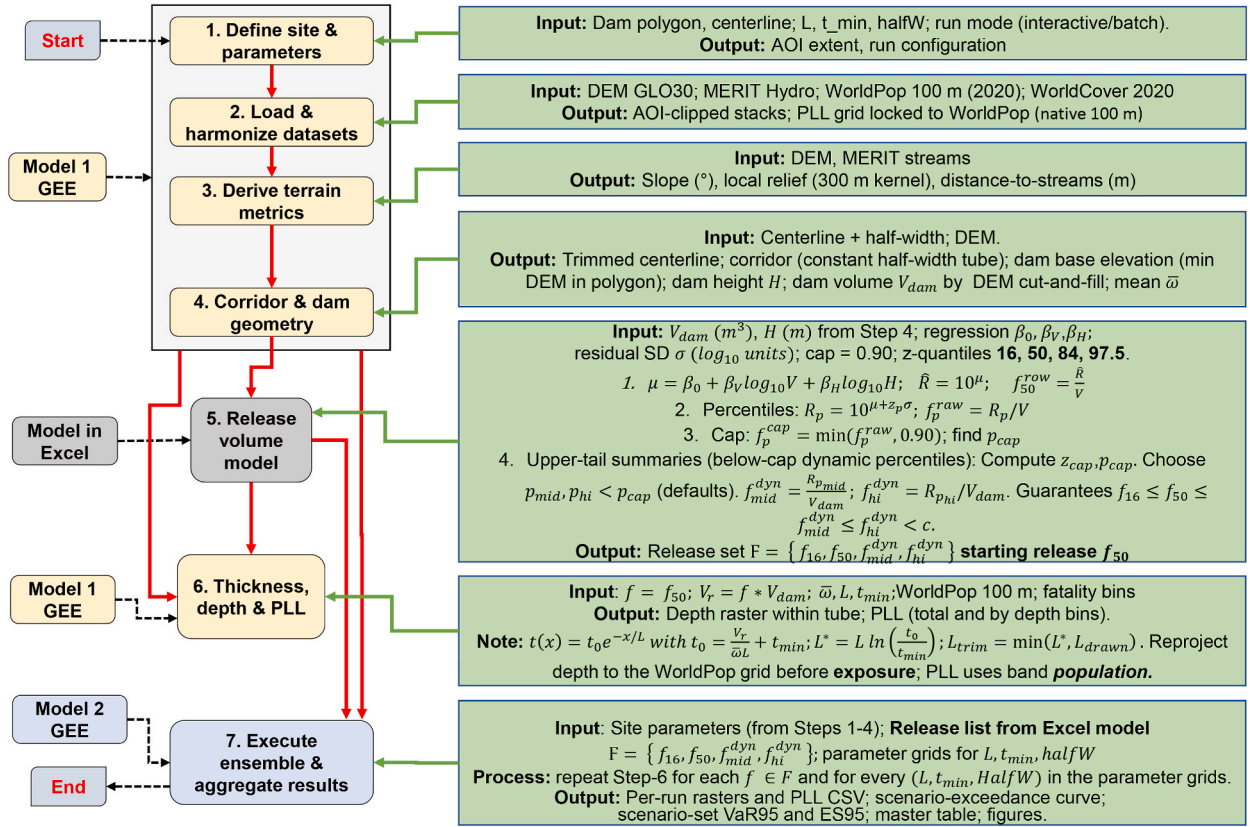


Fig. 1. Workflow for data-driven tailings-failure consequence screening.

spread σ and a physical cap at 0.90 of storage (Eq. (4)-(11b)) [19,20]. The interactive script uses f_{50} supplied by the Excel model for on-screen diagnostics. This is a screening-level consequence workflow, not a breach-hydraulics model. Throughout this paper, we use the term “framework” for the integrated workflow that combines DEM-based dam geometry, regression-based release estimates, corridor-based routing, PLL evaluation on WorldPop, and portfolio-level VaR95/ES95 diagnostics. The framework is designed to support portfolio-wide screening of conditional loss-of-life consequences across multiple tailings dams and to operate within, rather than replace, the site-specific breach and evacuation studies mandated by existing tailings-risk governance frameworks. All corridor and depth computations occur on the 30 m DEM grid. Before exposure, the depth raster is reprojected to the native WorldPop 100 m grid, and the population layer is never resampled.

For reproducibility, each run records a configuration comprising the dam polygon and a centerline guided by the Multi-Error-Removed Improved-Terrain (MERIT) Hydro dataset, the map projection and scale, the parameter set $\{L, t_{min}, halfW\}$, the selected release fraction(s) f_p capped at 0.90, and the script identifier, namely model-1-GEE for the interactive mode and model-2-GEE for the batch mode. These metadata are archived with every export.

Guided by MERIT-Hydro streams, the operator traces a centerline along the most plausible flow path. Using f_{50} together with user-set L (decay length), t_{min} (minimum meaningful thickness), and half-width, model-1-GEE builds a longitudinal exponential thickness field along the centerline (Eq. (13)), computes the theoretical stop length and trimming rules (Eqs. (15) and (16)), and buffers the trimmed centerline to form a constant half-width tube that serves as the baseline corridor. Thickness is computed on the terrain grid and then reprojected to the native WorldPop 100 m grid before exposure [23,24]. Potential Loss of Life (PLL) is evaluated on that grid using the population band and the depth-bin fatality curve reported in Table 4, with aggregation according to Eq. (18) [17,25]. A DEM-derived valley mask (distance-to-stream, slope, local relief) may be rendered for diagnostics, but it is not applied to constrain the baseline corridor.

The scenario-set run (model-2-GEE) repeats the thickness→depth→PLL pipeline for each $f \in F$ and for all combinations of L, t_{min} , and half-width in predefined parameter grids (full-factorial). Outputs include per-run depth rasters and PLL comma-separated-value (CSV) tables, together with site-level scenario-set summaries consisting of the scenario-exceedance curve, VaR95 and ES95. All datasets follow frozen versions and access dates (DEM GLO30, MERIT Hydro, WorldPop 2020, WorldCover 2020); projections, scales, parameters, and geometries are archived to ensure full reproducibility.

Fig. 1 summarises the sequence of operations implemented by model-1-GEE, the Excel release model and model-2-GEE; the individual steps are described in Sections 2.1–2.7.

2.1. Site inputs, portfolio definition, and area-of-interest construction - workflow step 1

The analysis is performed on a site-specific area of interest (AOI) defined by two user inputs: a dam polygon and a downstream centerline. At this step (executed in separate scripts), we set the model parameters $\{L(m), t_{min}(m), halfW(m)\}$. The interactive script (model-1-GEE) uses the median release f_{50} supplied by the Excel release model for on-screen diagnostics, whereas the batch script (model-2-GEE) ingests the full percentile set $\{f_{16}, f_{50}, f_{mid}^{dyn}, f_{hi}^{dyn}\}$ from the same Excel model and enumerates the parameter grids. The dam polygon delineates the impoundment and is used to derive terrain-based dam metrics on the native 30 m DEM grid [22]. We compute a reference base elevation z_{base} as the minimum DEM value inside the polygon (after AOI clipping). We define the crest elevation as the polygon-wise DEM maximum, and compute dam height as:

$$H = z_{crest} - z_{base} \quad (1)$$

where H is the DEM-derived dam height (m), z_{crest} is the maximum DEM elevation within the dam polygon (m), and z_{base} is the minimum DEM elevation within the same polygon (m).

The stored volume is obtained by DEM cut-and-fill above the base within the dam polygon:

$$V_{dam} = \int_{\text{geometry}} \max(\text{DEM}(x) - z_{base}, 0) dA \quad (2)$$

where V_{dam} is the estimated stored volume above the reference base elevation (m^3), geometry denotes the dam polygon, $\text{DEM}(x)$ is the terrain elevation at horizontal location $x(m)$, $\max[\text{DEM}(x) - z_{base}, 0]$ is the positive elevation difference above the base, and dA is an infinitesimal horizontal area element (m^2).

Discretising on the raster gives:

$$V_{dam} = \sum_{i \in \text{geometry}} \max(\text{DEM}_i - z_{base}, 0) A_{\text{pix},i} \quad (3)$$

where i indexes the DEM pixels within the dam polygon, DEM_i is the elevation of pixel $i(m)$, $A_{\text{pix},i}$ is its horizontal area obtained from `ee.Image.pixelArea` (m^2), and $\max(\text{DEM}_i - z_{base}, 0)$ is the positive thickness above the reference base (m). Their product is the volume represented by pixel i , and summation over all pixels yields V_{dam} [26].

We archive (V_{dam}, H) , together with DEM metadata (projection, resolution, access date) for reproducibility. This DEM-based construction of base elevation, crest elevation and dam height is illustrated schematically in Fig. 2(a). The geometric constructions for dam height, the plain-tube corridor, and the longitudinal thickness field are summarised in Fig. 2(a–c).

The downstream centerline and baseline corridor are constructed using MERIT Hydro guidance as described in Section 2.3.

2.2. Datasets and spatial standards - workflow step 2

We used global, open datasets and preserve each product's native resolution and projection until harmonisation is required by the computation. Topography comes from COPERNICUS/DEM/GLO30 (30 m) and provides terrain derivatives used downstream (slope and a 300 m local-relief metric) [22]. Hydrographic context was provided by Multi-Error-Removed Improved-Terrain (MERIT) Hydro v1.0.1 (MERIT/Hydro/v1_0_1; bands `upa` and `dir`), developed and documented by Yamazaki et al. [27], and used to guide the centerline along objectively derived drainage. Population exposure is evaluated with WorldPop/GP/100m/pop (2020) using the band

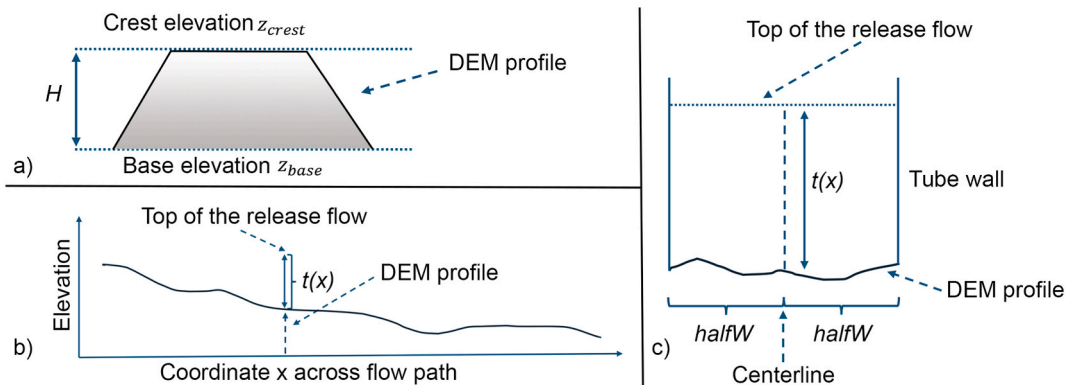


Fig. 2. Schematic dam height, longitudinal thickness and corridor cross-section. (a) Dam height H defined as the elevation difference between crest z_{crest} and base z_{base} in the DEM. (b) Longitudinal DEM profile along the flow path, showing thickness $t(x)$ from the DEM surface to the top of the release flow at distance x downstream. (c) Cross-section of the constant-width corridor, with half-width ($halfW$) on either side of the centerline and thickness $t(x)$ above the DEM.

population on its native 100 m grid [28]. European Space Agency (ESA) WorldCover 10 m 2020 v100 (ESA/WorldCover/v100/2020) was used for cartographic context only [29]. The year 2020 denotes the land-cover reference year, whereas 2021 is the dataset publication year. Core layers and roles are summarised in Table 2. All core rasters are clipped to the AOI prior to analysis. These AOI-clipped stacks are the working inputs for subsequent steps.

Harmonisation proceeds in two stages. First, the corridor construction and the thickness-to-depth computations are performed on the terrain grid (30 m) within the plain half-width tube around the trimmed centerline, which serves as the baseline corridor. A valley-bottom diagnostic can be rendered for quality assurance (QA). It is not used to constrain the baseline corridor. Second, before any exposure calculations, the final depth raster is reprojected to the native WorldPop 100 m grid so that depth and population share the same projection, scale, and extent. Population is never resampled. Only depth is resampled using a continuous-field method appropriate for physical surfaces. All exposure metrics and PLL are computed strictly on the WorldPop grid.

Version tracking and reproducibility are recorded explicitly. We log collection identifiers, versions, and dates of access for all layers: DEM COPERNICUS/DEM/GLO30 (band *DEM*), hydro MERIT/Hydro/v1_0_1 (bands *upa*, *dir*), population WorldPop/GP/100m/pop (2020) (band *population*), and land cover ESA/WorldCover/v100/2020 (band *Map*). Processing is performed in the GEE Code Editor. AOI geometries, parameter values, and projections are archived with the outputs.

2.3. Terrain metrics, corridor construction, and dam geometry - workflow steps 3 and 4

This subsection corresponds to workflow steps 3-4 in Fig. 1 and covers.

- terrain diagnostics used to guide routing and QA,
- the MERIT-guided centerline and plain-tube corridor, and
- dam height and volume derived from the DEM.

We derive terrain diagnostics on the native 30 m DEM to support routing and quality assurance [22,30,31]. Slope is computed in degrees, and local relief is computed with a 300 m moving window to emphasise valley floors. To guide routing, we render a drainage skeleton from MERIT Hydro (using the upstream-area and flow-direction bands) and compute distance to that skeleton [27]. These layers are co-registered to the DEM grid and used for visualisation and analyst QA. They do not constrain the baseline corridor in the main results.

Fig. 3(a-c) illustrates a typical model-1-GEE diagnostic view for one site, showing the mapped dam footprint, MERIT-derived drainage network with flow-direction arrows, AOI hillshade and WorldPop population layers, and the interactively chosen centerline used for visual QA of the routing.

The downstream centerline is then digitised along the most plausible flow path using the MERIT-derived drainage as guidance. This anchors both the location and the direction of the candidate pathway to an external hydrography product and reduces subjectivity. The centerline represents the dominant downstream drainage path rather than the full two-dimensional flow field. Reduced-order and semi-physical tailings-screening approaches have similarly used DEM-controlled routing and empirical volume-area relationships to delineate first-order impact zones where site-specific breach hydrographs and rheological properties are unavailable [32,33]. In the present framework, the line is constrained by the MERIT Hydro drainage network and visually checked against the DEM and satellite imagery before being archived with the run configuration. Nevertheless, some analyst dependence remains, particularly where drainage divides are poorly defined, multiple downstream branches are possible, or local infrastructure modifies the natural flow path.

The polyline is buffered to a constant half-width tube, which serves as the baseline hazard corridor. A cross-section of this constant-width tube, showing the half-width (*halfW*) on either side of the centerline and the thickness $t(x)$ above the DEM, is given in Fig. 2(c). The tube area together with the drawn centerline length defines the mean width, and the corridor geometry is stored with projection and units.

We compute mean width as $\bar{w} = \frac{\text{tube area}}{L_{\text{drawn}}}$ (m) and archive $\bar{w}$, the tube area, and L_{drawn} with projection and units. Dynamic trimming by the stop-length criterion is applied later in Step 6, once a release fraction has been selected.

The dam-height and stored-volume estimates derived in Section 2.1 are summarised for all 17 study sites in Table 3 and provide the inputs to the release-volume regression described in Section 2.4.

Table 2
Core datasets and roles.

Layer	GEE ID/bands	Native resolution	Role in workflow
Digital elevation model	COPERNICUS/DEM/GLO30/DEM	30 m	Terrain analysis, slope, local relief, dam cut-and-fill
Hydrography	MERIT/Hydro/v1_0_1/upa, dir	~90 m	Potential drainage lines and distance-to-streams; guidance for centerline
Population	WorldPop/GP/100m/pop/population (2020)	100 m	Exposure grid and PLL aggregation on native grid
Land cover (context)	ESA/WorldCover/v100/2020/Map	10 m	Cartographic context; not used in PLL calculations

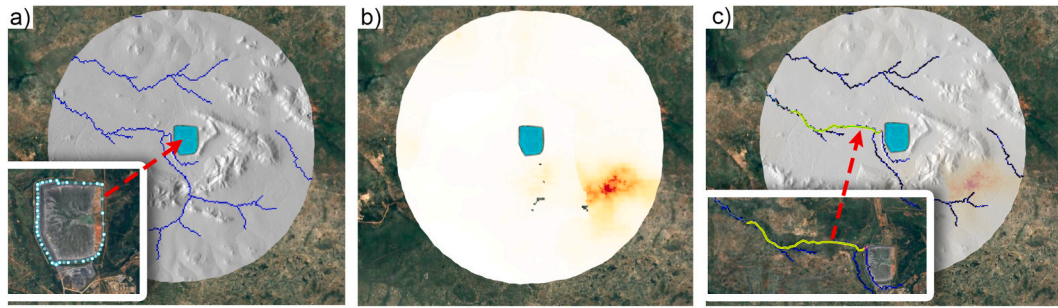


Fig. 3. Interactive diagnostics for site setup in the model-1-GEE script. (a) DEM hillshade around the circular AOI with MERIT-Hydro drainage network (blue) and mapped tailings dam polygon (cyan). Flow-direction arrows along the drainage network indicate plausible downstream routes. Inset: high-resolution satellite view of the digitised dam footprint. (b) WorldPop 100 m population within the same AOI, highlighting inhabited areas around the dam. (c) Example of an interactively chosen centerline (yellow) following a MERIT stream, which serves as the baseline flow path for the scenario set. Inset: satellite zoom on the selected stream reach.

2.4. Release-volume model (Excel) - workflow step 5

This step converts the dam metrics from Step 4 into a release-fraction set that can be injected into the GEE models. The inputs are stored volume V_{dam} (m^3) and dam height H (m). We fit a regression using the historical tailings-failure dataset reported by Picicullo et al. [6], Table 2). Prior studies have used similar regression-based approaches to estimate released volume/fraction from failure attributes and to quantify coefficient uncertainty [19,20]. The fitted model returns coefficients $\beta_0, \beta_V, \beta_H$ and a residual spread σ in $\log_{10}$ units. We evaluate percentiles at $p \in \{16, 50, 84, 97.5\}$ and apply a physical cap $c = 0.90$ (fraction of storage). For upper-tail reporting we use below-cap probabilities $p_{mid}, p_{hi} < p_{cap}$ (Eq. (10)–(11b)). Standard-normal z-quantiles used: $z_{16} = -0.9945, z_{50} = 0, z_{84} = +0.9945, z_{97.5} = +1.9600$.

Units: R_p is a volume (m^3); $f_p = R_p/V$ and the cap 0.90 are dimensionless.

The regression model is:

$$\mu = \beta_0 + \beta_V \log_{10}(V_{dam}) + \beta_H \log_{10}(H) \quad (4)$$

where μ is the predicted mean of $\log_{10} R$, with R denoting released volume in m^3 ; β_0 is the regression intercept; β_V and β_H are the coefficients for stored volume and dam height, respectively; V_{dam} is stored volume (m^3); and H is dam height (m).

Percentiles of the released volume R map via Eq. (5). The (raw) release fraction f_p^{row} follows Eq. (6), with the physical cap in Eq. (7). To avoid cap-induced pile-ups in the reported upper tail, we summarise it using below-cap dynamic percentiles per Eqs. (10)–(11b), rather than smoothing capped values.

For model transparency, the Excel ‘Config’ sheet reports the regression fit based on the compiled historical-failures dataset in the Data sheet, comprising $N = 54$ events and 51 residual degrees of freedom ($df = 51$). Coefficients are estimated in $\log_{10}$ space and reported with standard errors (SE) and 95% confidence intervals (CIs) ($\hat{\beta} \pm t_{0.975,51} \cdot SE, t \approx 2.01$): $\beta_0 = -0.0476$ (SE 0.5881; 95% CI $[-1.230, 1.135]$), $\beta_V = 0.7049$ (SE 0.1149; 95% CI $[0.474, 0.936]$), $\beta_H = 0.8543$ (SE 0.3210; 95% CI $[0.209, 1.500]$); residual standard

Table 3

Basic characteristics of the 17 tailings dams considered in this study, including estimated dam height and stored tailings volume derived from the DEM-based cut-and-fill procedure.

Site	Country	Estimated tailings dam height H (m)	Estimated stored tailings volume V_{dam} (m^3)
Loulo-Goukoto	Mali	19	19,476,373
Geita	Tanzania	55	104,733,620
KCGM	Australia	26	55,598,525
South Deep	South Africa	38	56,678,380
Fosterville	Australia	35	7,496,855
Mponeng	South Africa	39	26,133,714
Kuranakh	Russia	37	78,131,021
Driefontein	South Africa	42	38,939,213
Sigui	Guinea	52	83,258,182
Young-Davidson	Canada	57	17,598,689
Edikan	Ghana	57	69,275,453
Haile	United States	41	39,578,850
Ergo	South Africa	83	491,457,833
Goldex	Canada	8	2,474,817
Chirano	Ghana	61	23,912,462
Aurizona	Brazil	35	7,900,615
Yanfolila	Mali	23	8,369,300

Table 4
Default depth-fatality mapping $\pi(h)$

Water/flow depth $h(m)$	$\pi(h)$
$h < 0.3$	0
$0.3 \leq h < 1$	0.1
$1 \leq h < 2$	0.2
$2 \leq h < 5$	0.5
$h \geq 5$	0.9

deviation (SD) $\sigma \log_{10} = 0.5694$ (i.e., a one-sigma multiplicative scatter of $10^{0.5694} \approx 3.71$ in R); $R^2 = 0.6465$.

Interpretation: holding the other predictor fixed, a $10\times$ increase in stored volume multiplies the predicted release by $10^{\beta_V} \approx 5.07$, and a $10\times$ increase in dam height multiplies it by $10^{\beta_H} \approx 7.15$; β_V and β_H are statistically different from zero (CIs exclude 0), while the intercept is not.

For a target percentile p with standard-normal deviate z_p , the release volume and the corresponding raw fraction are:

$$R_p = 10^{\mu + z_p \sigma} [\text{m}^3] \quad (5)$$

where R_p is the released-volume estimate at percentile probability p (m^3), $z_p = \Phi^{-1}(p)$ is the corresponding standard-normal deviate, $\Phi^{-1}(\cdot)$ is the inverse standard-normal cumulative distribution function, and $\sigma_{\log_{10}}$ is the residual standard deviation of the regression in $\log_{10} (\text{m}^3)$ units.

$$f_p^{\text{raw}} = \frac{R_p}{V_{\text{dam}}} \quad (6)$$

where f_p^{raw} is the uncapped released fraction at percentile probability p , calculated as the ratio between R_p and V_{dam} .

We enforce the physical bound by capping the fraction at $c = 0.90$ [20]:

$$f_p^{\text{cap}} = \min(f_p^{\text{raw}}, c) \quad (7)$$

where f_p^{cap} is the capped released fraction and $c = 0.90$ is the maximum permitted fraction of stored volume released.

Capping all high percentiles can otherwise produce artificial pile-ups at 0.90. To avoid this, we compute the percentile at which the cap first binds:

$$z_{\text{cap}} = \frac{\log_{10}(cV_{\text{dam}}) - \mu}{\sigma \log_{10}} \quad (8)$$

$$p_{\text{cap}} = \Phi(z_{\text{cap}}) \quad (9)$$

where z_{cap} is the standard-normal deviate at which the uncapped release reaches cV_{dam} , p_{cap} is the corresponding cumulative probability, and $\Phi(\cdot)$ is the standard-normal cumulative distribution function.

Instead of smoothing capped values, we report two below-cap probabilities p_{mid} and p_{hi} chosen strictly below p_{cap} . These produce uncapped, ordered fractions that remain $< c$ without altering sub-cap percentiles:

$$0 < p_{\text{mid}} < p_{\text{hi}} < p_{\text{cap}} \quad (10)$$

$$f_{\text{mid}}^{\text{dyn}} = \frac{R_{p_{\text{mid}}}}{V_{\text{dam}}} \quad (11a)$$

$$f_{\text{hi}}^{\text{dyn}} = \frac{R_{p_{\text{hi}}}}{V_{\text{dam}}} \quad (11b)$$

Here, p_{mid} and p_{hi} are two selected percentile probabilities satisfying $0 < p_{\text{mid}} < p_{\text{hi}} < p_{\text{cap}}$. The quantities $R_{p_{\text{mid}}}$ and $R_{p_{\text{hi}}}$ are obtained from Eq. (5), while $f_{\text{mid}}^{\text{dyn}}$ and $f_{\text{hi}}^{\text{dyn}}$ are the corresponding below-cap release fractions. The final release set is $F = \{f_{16}, f_{50}, f_{\text{mid}}^{\text{dyn}}, f_{\text{hi}}^{\text{dyn}}\}$, where f_{16} and f_{50} denote the 16th-percentile and median release fractions, respectively.

This selection guarantees $f_{16} \leq f_{50} \leq f_{\text{mid}}^{\text{dyn}} \leq f_{\text{hi}}^{\text{dyn}} < 0.90$ and removes cap-induced plateaus. The release set is $F = \{f_{16}, f_{50}, f_{\text{mid}}^{\text{dyn}}, f_{\text{hi}}^{\text{dyn}}\}$. The interactive run (model-1-GEE) uses f_{50} to produce an on-screen PLL estimate, while the batch script (model-2-GEE) ingests the full set F and evaluates all combinations across the parameter grids of L , t_{min} and halfW . For reproducibility, the workbook stores auxiliary diagnostics (e.g., R_{50} , σ , z_{cap} , p_{cap}) and writes a per-dam CSV with fields: site_id , $V_{\text{dam}} (\text{m}^3)$, $H(m)$, f_{16} , f_{50} , $f_{\text{mid}}^{\text{dyn}}$, $f_{\text{hi}}^{\text{dyn}}$.

To evaluate the out-of-sample predictive performance of the release-volume regression, we conducted leave-one-out cross-validation across the 54 historical failure records. In each iteration, one event was excluded, the log-space regression was refitted using the

remaining 53 events, and the released volume of the omitted event was predicted without using that event during model estimation. Validation performance was assessed using the cross-validated coefficient of determination, root-mean-square error (RMSE), mean absolute error (MAE), and mean prediction error in $\log_{10}(\text{m}^3)$ units. We also evaluated the empirical coverage of fold-specific prediction intervals based on the residual standard deviation estimated from each training subset. This procedure provides an out-of-sample assessment of the regression component, although it does not by itself validate the downstream-routing or PLL components of the complete framework.

The historical-failure database may also be affected by reporting and selection bias because large, damaging or well-investigated events are more likely to be documented than smaller, remote or poorly recorded failures. The dataset does not provide event-specific inclusion probabilities, and the number and characteristics of undocumented failures are unknown. Consequently, a formal correction using inverse-probability weighting, models accounting for systematically unrecorded events, or related methods cannot be supported without introducing additional assumptions. The regression is therefore treated as conditional on the available documented-event sample and is used to generate release scenarios rather than as an unbiased estimate of the complete population of TSF failures. Residual-based percentiles, the physical release cap and leave-one-out cross-validation represent uncertainty and predictive variability within the recorded sample, but they do not remove reporting bias. Preferential recording of larger failures may shift the fitted relationship toward more severe documented events, particularly when applied to smaller or underrepresented facilities.

2.5. Thickness, depth, and potential loss of life (PLL) - workflow step 6

We convert a selected release fraction into a longitudinal thickness field, rasterise depth inside the plain half-width tube, and compute PLL on the native WorldPop 100 m grid. We select a release fraction $f \in F$ from Step 5 (interactive f_{50} ; batch: all $f \in \{f_{16}, f_{50}, f_{mid}^{dyn}, f_{hi}^{dyn}\}$) and compute released volume [34]:

$$V_r = f V_{dam} (\text{m}^3) \quad (12)$$

where V_r is the released volume used in the routing calculation (m^3), $f \in F$ is the selected released fraction, and V_{dam} is the estimated stored volume (m^3).

The corridor geometry supplies the mean width $\bar{w}$, while the user specifies the e-folding length L and the minimum meaningful thickness t_{min} . The exponential profile is used as a reduced-order parameterisation that produces monotonic downstream thinning and a finite stop length when the minimum-thickness criterion is reached. It is not presented as a universal rheological law for tailings flows. Similarly, the constant-width tube is a scenario envelope rather than a prediction of the local inundation boundary. Its half-width is varied across the scenario set, together with the decay length and minimum-thickness threshold, to represent uncertainty in lateral extent and longitudinal attenuation. Detailed numerical back-analyses demonstrate that actual tailings-flow depth and extent depend on the breach hydrograph, released volume, terrain and non-Newtonian rheology, and that different parameter combinations may reproduce similar bulk runout behaviour [7,35]. Accordingly, $t(x)$ is interpreted here as an effective screening thickness used to classify exposure, rather than a site-specific prediction of local hydrodynamic depth.

Thickness decays exponentially along the MERIT-guided centerline with chainage.

x measured from the dam toe:

$$t(x) = t_0 e^{-x/L} \quad (13)$$

$$t_0 = \frac{V_r}{\bar{w}L} + t_{min} \quad (14)$$

where $t(x)$ is the effective screening thickness at downstream chainage $x(m)$, t_0 is the intercept thickness at $x = 0 (m)$, x is distance along the centerline measured from the dam toe (m), L is the e-folding decay length (m), $\bar{w}$ is the mean corridor width (m), and t_{min} is the minimum meaningful thickness used as the stopping threshold (m).

When $t(x)$ reaches t_{min} the flow is assumed to stop. The corresponding stop length and applied trimming are:

$$L^* = L \ln\left(\frac{t_0}{t_{min}}\right) \quad (15)$$

$$L_{trim} = \min(L^*, L_{drawn}) \quad (16)$$

where L^* is the theoretical stop length (m), defined by $t(L^*) = t_{min}$; L_{drawn} is the available digitised centerline length (m); and L_{trim} is the applied corridor length (m), defined as the shorter of L^* and L_{drawn} .

The resulting longitudinal thickness field $t(x)$ along the flow path, measured as distance x downstream from the dam, is sketched in Fig. 2(b). When the digitised centerline extends to the theoretical stopping distance ($L_{drawn} \geq L^*$), the complete thickness profile conserves the released volume because [34]:

$$\bar{w} \int_0^{L^*} t(x) dx = \bar{w}L(t_0 - t_{min}) = V_r \quad (17)$$

The integral is evaluated from the dam toe, $x = 0$, to the theoretical stopping distance, $x = L^*$. Under the constant mean-width assumption, the complete theoretical profile integrates to V_r . If the available centerline is shorter than L^* , the calculations are performed over $L_{trim} = L_{drawn}$, and only the portion of the theoretical volume within the available corridor is represented.

We evaluate $t(x)$ along the polyline, paint the values to pixels inside the trimmed tube, and set depth to zero outside.

For exposure, the effective screening-thickness raster, computed on the 30 m terrain grid, is used as an inundation-depth proxy and reprojected to the native WorldPop 100 m grid before any counts so that depth and population share identical projection, scale, and extent [23]. Population is never resampled, only depth is resampled using a continuous-field method suitable for physical surfaces. All exposure metrics and PLL are therefore computed strictly on the WorldPop grid [23].

Fatality probability is linked to depth by a step function $\pi(h)$ whose default depth bins and fatality probabilities are listed in Table 4 [17,36].

Expected fatalities (PLL) are summed on the WorldPop 100 m grid as:

$$PLL = \sum_{i \in \Omega} \pi(d_i) \text{ population}_i \quad (18)$$

where PLL is conditional Potential Loss of Life, expressed as expected fatalities; Ω is the set of WorldPop cells intersecting the trimmed corridor; i indexes those cells; d_i is the reprojected effective depth proxy in cell i ; $\pi(d_i)$ is the fatality probability assigned to that depth; and population_i is the WorldPop population count in cell i . Optional by-bin summaries use bins B_k with probabilities π_k :

$$P_k = \sum_{i \in \Omega: d_i \in B_k} \text{population}_i \quad (19)$$

$$PLL_k = \pi_k P_k \quad (20)$$

where B_k is depth bin k , π_k is the fatality probability assigned to that bin, P_k is the population exposed within bin k , and PLL_k is the corresponding expected-fatality contribution. Summing P_k over all bins gives the total exposed population, while summing PLL_k gives the total conditional PLL in Eq. (18).

This calculation is applied to every parameter combination evaluated in Section 2.6. To examine uncertainty in the transferable fatality relationship, we performed a warning-response sensitivity analysis for the five sites classified as high consequence in Section 3.1. The current depth-fatality mapping was retained as the reference variant, without explicit credit for warning or evacuation. Two additional variants were derived from the Brumadinho Life Safety Model results reported by Lumbroso et al. [7]. In that study, modelled fatalities decreased from 456 without warning to 299-354 when warning was issued at the time of failure, giving a response multiplier of 0.656-0.776. A 5 min advance warning reduced fatalities by 200-270, leaving 186-256 fatalities and giving a multiplier of 0.408-0.561. For each scenario-set run i , the sensitivity-adjusted PLL was calculated as:

$$PLL_{i,v} = m_v PLL_{i,ref} \quad (21)$$

where $PLL_{i,ref}$ is the reference conditional PLL, m_v is the warning-response multiplier, and $PLL_{i,v}$ is the adjusted PLL for vulnerability variant v . The multipliers were applied in post-processing to the 144 run-level PLL outcomes for each high-consequence site, after which Var95 and ES95 were recalculated. These variants are illustrative transfer-based sensitivity bounds, not site-specific predictions of warning effectiveness. Building stock, warning dissemination, evacuation capacity and temporal occupancy remain outside the screening model.

2.6. Scenario-set execution and aggregation - workflow step 7

Model-2-GEE applies the thickness, reprojection and PLL procedure defined in Section 2.5 to the complete Cartesian product of the release set F , e-folding lengths L , minimum-thickness thresholds t_{min} and corridor half-widths.

Table 5

Model-2-GEE parameter grids used to generate the deterministic scenario set (144 runs per site).

Parameter	Symbol	Values/definition	Units	Levels	Screening-range basis
Released fraction set (cap-aware; regression-based)	F	$F = \{f_{16}, f_{50}, f_{mid}^{dyn}, f_{hi}^{dyn}\}$ where $f_{16} = P16$ and $f_{50} = P50$; f_{mid}^{dyn} and f_{hi}^{dyn} are site-specific upper-tail members (computed from the capped regression distribution)	-	4	Site-specific central and upper-tail regression scenarios, capped at 0.90 of storage. Not prior probabilities.
Longitudinal decay length	L	1000; 1500; 2000	m	3	Brackets rapid to gradual downstream attenuation and associated stopping lengths.
Minimum meaningful thickness (stopping threshold)	t_{min}	0.05; 0.10; 0.15	m	3	Tests retention or trimming of shallow distal deposits. All values are below the 0.30 m first non-zero fatality threshold.
Corridor half-width	$halfW$	40; 60; 80; 120	m	4	Total widths of 80-240 m span narrow, confined corridors to broader spreading envelopes at 30 m DEM resolution.

The parameter levels in Table 5 were selected as transparent screening bounds rather than as quantiles of calibrated prior distributions. The e-folding lengths $L = 1000, 1500$ and 2000 m provide a factor-of-two bracket from comparatively rapid to gradual longitudinal attenuation. Together with released volume, corridor width and t_{min} , these values control the theoretical stopping length. The minimum-thickness values $t_{min} = 0.05, 0.10$ and 0.15 m test whether shallow distal deposits are retained or removed. All three lie below the first non-zero fatality-depth threshold of 0.30 m in the adopted mapping. Corridor half-widths of $40, 60, 80$ and 120 m correspond to total widths of $80\text{--}240$ m, spanning a relatively narrow valley-confined tube to a broader lateral-spread envelope while remaining spatially resolved by the 30 m terrain grid. These levels are not treated as universal physical limits or prior probabilities. They define a reproducible design space for comparative sensitivity analysis. The four release-fraction scenarios, three values of L , three values of t_{min} and four corridor half-widths produce $4 \times 3 \times 3 \times 4 = 144$ combinations per site. These combinations constitute a full-factorial design for structured parameter-sensitivity analysis within the selected screening bounds. They are not random draws from a calibrated joint probability distribution. Each combination is assigned an equal design weight of $1/144$ so that every selected parameter level is represented transparently and consistently across all facilities. Equal weighting therefore does not imply that the physical scenarios are equally likely.

Probabilistic Monte Carlo and Latin hypercube sampling were considered. However, these approaches require defensible marginal distributions and, where relevant, dependence relationships among the uncertain inputs [37]. Such distributions are not currently available for L , t_{min} , corridor half-width or their site-specific dependence on release magnitude. Imposing common probability distributions across all 17 facilities would therefore introduce assumptions that cannot presently be supported by data. The equal-weight full-factorial design was selected as a reproducible and assumption-transparent baseline. When calibrated site-specific distributions or expert-elicited weights become available, the archived run-level outputs can be reweighted without rerunning the geospatial calculations. Runs producing a zero-length or empty tube yield zero exposure and remain in the design so that all enumerated parameter combinations are retained.

The full Cartesian product does not impose statistical independence in a probabilistic sense because no joint probability distribution is fitted. Instead, it enumerates every release-routing combination to expose main effects and interactions without imposing an unsupported dependence structure. However, the equal-weight baseline does not represent possible physical dependence among the inputs. To test sensitivity to the possible association between released fraction and lateral extent, two bounding rank-coupling configurations were evaluated using the archived run-level outcomes. In the positive-coupling configuration, the four release fractions, ordered from smallest to largest, were paired with corridor half-widths of $40, 60, 80$ and 120 m, respectively. In the opposite bounding configuration, the ordered release fractions were paired with half-widths of $120, 80, 60$ and 40 m. All three L levels and all three t_{min} levels were retained, producing 36 equally design-weighted combinations in each dependence stress test. VaR95, ES95 and cross-site rankings were recalculated for both configurations. These stress tests are dependence bounds, not calibrated joint probability models.

From the scenario set, we obtain a vector of 144 PLL totals per facility, expressed as expected fatalities. Each value corresponds to one enumerated parameter combination and receives the equal design weight $1/144$. The resulting distribution is an empirical design distribution rather than a calibrated probability distribution. It is summarised using a scenario-exceedance curve, scenario-set VaR95 and scenario-set ES95.

- Scenario-exceedance curve [38]

For a PLL threshold q , the fraction of enumerated parameter combinations whose PLL exceeds that threshold is:

$$SE(q) = \frac{1}{N} \sum_{i=1}^N 1(PLL_i > q) \quad (22)$$

where $SE(q)$ is the scenario-exceedance fraction at PLL threshold q , q is expressed in expected fatalities, $N = 144$ is the number of enumerated parameter combinations, PLL_i is the PLL outcome of combination i , and $1(\cdot)$ is the indicator function, equal to 1 when $PLL_i > q$ and 0 otherwise. The scenario-exceedance curve is the complement of the empirical cumulative distribution function under the equal-weight design and is plotted as a right-continuous step function. It describes the proportion of enumerated parameter combinations exceeding a specified PLL value. It is not the probability of dam failure, the annual probability of a loss-of-life outcome or a calibrated conditional probability given failure.

- Scenario-set Value-at-Risk at 95% (VaR95) [39]

The scenario-set VaR95 is calculated as the 95th percentile of the outcome vector y using the Hyndman-Fan Type-7 empirical quantile [40]:

$$VaR_{95} = Q_{0.95}(y) \quad (23)$$

where $y = (y_1, \dots, y_N)$ is the vector of PLL outcomes from the $N = 144$ parameter combinations and $Q_{0.95}(y)$ is the 95th-percentile operator calculated using the Hyndman-Fan Type-7 empirical quantile definition.

Interpretation: approximately 5% of the enumerated, equally design-weighted parameter combinations exceed this value. The 5% refers only to the scenario set and must not be interpreted as a physical or annual exceedance probability.

- Scenario-set Expected Shortfall at 95% (ES95) [39]

In this study, ES95 is calculated as the empirical mean of the scenario outcomes in the upper tail defined by the VaR95 threshold:

$$T_{95} = \{i : y_i \geq \text{VaR}_{95}\} \quad (24)$$

$$\text{ES}_{95} = \frac{1}{|T_{95}|} \sum_{i \in T_{95}} y_i \quad (25)$$

where $T_{95} = \{i : y_i \geq \text{VaR}_{95}\}$ is the set of run indices belonging to the upper tail, $|T_{95}|$ is the number of runs in that set, y_i is the PLL outcome of run i , and ES95 is their arithmetic mean. Because the design is finite and deterministic, $|T_{95}|$ may differ slightly from exactly 5% of N , particularly when several outcomes equal the VaR95 threshold. ES95 is therefore interpreted as the threshold-defined upper-tail mean of the equal-weight scenario set. It does not represent an annualised expected loss or a probability-weighted expectation based on calibrated physical likelihoods. Units: expected fatalities.

Together with the 50th, 90th and 99th percentiles (P50, P90 and P99), these statistics summarise central tendency and upper-tail behaviour. The resulting tables, summary metrics and optional rasters are described in Section 2.7.

These outputs provide the quantitative basis for the Results and Discussion (Section 3). For each of the 17 tailings dams, we use the empirical scenario-set PLL distribution and its summaries (P50, P90, P99, VaR95 and ES95) to construct four complementary diagnostics: percentile “league tables,” scenario-exceedance curves, VaR95-ES95 scatterplots and PLL distribution boxplots. In Section 3.1–3.5 these diagnostics are used to compare conditional consequences across the portfolio and to group sites into high-, moderate- and low-consequence classes based on their PLL metrics. This three-tier grouping is a study-specific construct derived from the scenario set and is not intended to replace formal consequence categories in existing standards.

To examine the relative cross-site associations of tail PLL severity with downstream exposure and facility size, we calculated pairwise and partial Spearman rank correlations across the 17 sites. The variables were scenario-set ES95, estimated stored tailings volume and a standardised reference-run exposure measure. For each facility, exposed population was defined as the [28] population within the trimmed corridor generated using the site-specific median release fraction f_{50} , $L = 1500$ m, $t_{\min} = 0.10$ m and a corridor half-width of 60 m. Two partial associations were evaluated: ES95 with exposed population while controlling for stored volume, and ES95 with stored volume while controlling for exposed population. Partial correlations were calculated by residualising the ranked variables against the ranked control variable and correlating the resulting residuals. The rank-based formulation reduces sensitivity to the several-orders-of-magnitude ranges and does not assume linear relationships on the original scale. Because the analysis includes only 17 facilities, it was treated as exploratory. Robustness to individual observations was assessed by recalculating the partial correlations after omitting each site in turn. Two-sided significance was evaluated using residual-permutation tests with 200,000 permutations.

For auxiliary context, Appendix A lists the original disclosure-based consequence or hazard classes reported by Franks et al. [10] alongside the study-specific conditional PLL tiers. The original class labels and classification-system identifiers are retained without harmonisation because the underlying schemes differ in terminology, criteria and consequence dimensions, and may include environmental, economic, infrastructure or other impacts in addition to potential loss of life. No agreement score or equivalence between the two classifications is inferred, and the appendix comparison is not used for model validation, study-tier assignment or Stage 2 escalation decisions. Independent quantitative evaluation of the release-volume component was instead performed using leave-one-out cross-validation of the 54 historical failure records (Sections 2.4 and 3.7; Table 9). The Brumadinho case is presented separately as a screening-level plausibility check of the selected downstream route, applied corridor extent and workforce-exposure

Table 6

Summary of conditional PLL metrics (expected fatalities) for the 17 tailings dams: median (P50), upper percentiles (P90, P99), Value-at-Risk at 95% (VaR95 = P95), and Expected Shortfall at 95% (ES95).

Site	Consequence group (High/Moderate/Low)	P50	P90	P99	VaR95 (P95)	ES95
Loulo-Goukoto	High	15.0	63.6	105.66	82.8	90.5
Geita	High	28.7	61.8	91.02	71.3	82.6
KCGM	Moderate	0.1	10.2	46.02	36.7	40.4
South Deep	Low	1.7	4.4	4.77	4.5	4.6
Fosterville	Low	0.2	0.4	0.67	0.5	0.6
Mponeng	Low	1.3	2.7	3.82	3.6	3.7
Kuranakh	Moderate	4.3	17.3	34.75	22.2	28.4
Driefontein	Moderate	0.1	0.6	60.29	0.7	22.6
Siguiri	High	43.0	101.4	196.87	132.3	157.2
Young-Davidson	Low	0.1	0.6	1.40	0.8	1.0
Edikan	High	36.2	76.3	140.12	93.5	115.9
Haile	Low	0.3	1.1	1.41	1.2	1.3
Ergo	Low	1.8	5.1	8.49	6.3	7.1
Goldex	Low	0.0	0.1	0.15	0.1	0.1
Chirano	High	59.8	129.4	185.95	148.1	165.3
Aurizona	Low	0.1	0.2	0.34	0.2	0.3
Yanfolila	Low	0.4	3.7	6.69	5.1	6.0

assumptions (Section 3.7; Fig. 9; Table 10).

2.7. Implementation details and exports

All geospatial processing was carried out in the Google Earth Engine Code Editor using model-1-GEE for interactive site setup and model-2-GEE for deterministic batch execution. The release-volume regression and associated diagnostics were implemented in the accompanying Excel workbook. The scenario design is deterministic and requires no random seeds for the GEE runs. Fixed random seeds were used and archived for the residual-permutation tests in the cross-site correlation analysis. Runs producing a zero-length or empty corridor are retained with zero exposure so that all 144 enumerated parameter combinations remain represented. We export the following.

1. 30 m terrain derivatives (slope, local relief),
2. the trimmed centerline and the tube corridor as vectors,
3. optional per-run depth rasters (30 m) and their WorldPop-grid versions (100 m) for mapping, and
4. CSV tables: a per-run master table with parameters and diagnostics ($f, L, t_{min}, halfW, \bar{\omega}, H, V_{dam}, V_r, L_{trim}, PLL, PLL$ by bin), plus site-level summaries containing the scenario-exceedance curve and tail metrics (VaR95, ES95).

All exports carry projection, scale, units, and timestamps. Dataset identifiers, collection versions, and dates of access are logged in the scripts to ensure exact reproducibility.

3. Results and Discussion

In this section we present the consequences of hypothetical tailings dam failures across the 17-site portfolio, expressed as conditional potential loss of life (PLL) under the scenario set defined in Section 2. We first summarise site-level PLL metrics and group the dams into three consequence classes (Section 3.1), then examine their ranking, scenario-exceedance behaviour, tail severity and within-site variability (Sections 3.2–3.5). We then examine how tail PLL severity relates to estimated stored volume and downstream exposed population (Section 3.6), before synthesising the findings and their implications for tailings governance (Section 3.7).

3.1. Overview of scenario-set PLL outcomes and site grouping

For each dam, model-2-GEE takes the four cap-aware release fractions $F = \{f_{16}, f_{50}, f_{mid}^{dyn}, f_{hi}^{dyn}\}$ obtained from the regression-based release model and evaluates them over a pre-defined deterministic grid of corridor-thickness parameters $\{L, t_{min}, halfW\}$, yielding 144 scenario runs per site (Table 5). All 144 combinations are run through the pipeline that converts thickness to depth and then to PLL on the WorldPop grid, producing one total PLL value (expected fatalities) per parameter combination.

The resulting empirical design distribution of PLL for each site is summarised by percentiles (P50, P90 and P99), a scenario-exceedance curve and two scenario-set tail metrics: Value-at-Risk at 95% (VaR95, the 95th percentile of PLL) and Expected Shortfall at 95% (ES95, the empirical mean of scenario outcomes in the upper tail defined by the VaR95 threshold). ES95 should therefore be read as a scenario-set upper-tail mean conditional on failure, not as an annualised expected loss. All reported PLL, percentile, VaR95 and ES95 values describe consequences given that a dam failure occurs and given our scenario set of plausible parameter combinations. They therefore quantify conditional loss-of-life consequences, not annual failure risk. The equal design weights support transparent comparison among facilities but do not represent estimates that the 144 parameter combinations are physically equally likely. The resulting site-level PLL summaries are reported in Table 6.

Key PLL metrics for all 17 tailings dams (P50, P90, P99, VaR95 and ES95) are summarised in Table 6 and form the basis for the consequence grouping discussed below. Across the 17 study sites the scenario set outcomes span almost three orders of magnitude in

Table 7

Sensitivity of scenario-set VaR95 and ES95 at the five high-consequence sites to warning-response multipliers derived from the Brumadinho Life Safety Model.

Site	Reference VaR95	VaR95: warning at failure	VaR95: 5 min advance warning	Reference ES95	ES95: warning at failure	ES95: 5 min advance warning
Loulo-Goukoto	82.8	54.3-64.3	33.8-46.5	90.5	59.4-70.2	36.9-50.8
Geita	71.3	46.8-55.3	29.1-40.0	82.6	54.2-64.1	33.7-46.3
Siguiiri	132.3	86.8-102.7	54.0-74.2	157.2	103.1-122.0	64.1-88.2
Edikan	93.5	61.3-72.6	38.1-52.5	115.9	76.0-89.9	47.3-65.0
Chirano	148.1	97.2-114.9	60.4-83.1	165.3	108.4-128.3	67.4-92.7

Note: All values are expected fatalities. Reference VaR95 and ES95 values are taken from the baseline scenario-set results in Table 6. Warning-at-failure ranges were calculated using multipliers of 0.656-0.776, corresponding to reductions of approximately 22-34%. Five-minute advance-warning ranges were calculated using multipliers of 0.408-0.561, corresponding to reductions of approximately 44-59%. The multipliers were derived from the Brumadinho Life Safety Model results reported by Lumbruso et al. [7]. They represent transfer-based sensitivity bounds rather than calibrated site-specific estimates of warning effectiveness.

Table 8

Pairwise and partial Spearman associations of ES95 with standardised reference-run exposed population and estimated stored tailings volume across the 17-site portfolio.

ES95 relationship	Pairwise Spearman ρ	Control variable	Partial Spearman ρ	Permutation p – value	Leave-one-site-out partial- ρ range
Exposed population	0.78	Stored volume	0.64	0.006	0.48-0.79
Stored volume	0.60	Exposed population	0.16	0.529	0.04-0.33

Note: Partial correlations were calculated from ranked variables. Two-sided p -values were obtained from residual-permutation tests with 200,000 permutations. Leave-one-site-out ranges were calculated by omitting each facility in turn. The results are exploratory because the portfolio contains only 17 sites.

Table 9

Leave-one-out cross-validation performance of the historical release-volume regression.

Metric	Cross-validation result
Historical failures, N	54
Cross-validated R^2	0.607
RMSE, $\log_{10} (\text{m}^3)$	0.583
MAE, $\log_{10} (\text{m}^3)$	0.452
Mean error, $\log_{10} (\text{m}^3)$	0.005
Coverage of nominal 16th–84th percentile interval	75.9%
Coverage of nominal 2.5th–97.5th percentile interval	92.6%

Note: RMSE and MAE values of 0.583 and 0.452 in $\log_{10}$ space correspond to approximate multiplicative factors of 3.83 and 2.83, respectively. Mean error is defined as observed minus predicted $\log_{10}$ released volume. This analysis provides out-of-sample validation of the release-volume regression component only. It does not constitute end-to-end validation of the corridor geometry, effective-thickness representation, population exposure or PLL estimates.

Table 10

Brumadinho (2019) screening-level plausibility check: applied corridor extent and sensitivity of conditional PLL to workforce spatial distribution with released volume fixed at $9.7 \times 10^6 \text{ m}^3$.

Metric	Scenario B ₁ : workers spread (wide polygon)	Scenario B ₂ : workers concentrated (smaller polygon)
$V_r (\text{m}^3)$	9,700,000	9,700,000
$L_{\text{trim}} (\text{m})$	9028	9028
Population in corridor, WorldPop-only (Pop_A)	49	49
Exposed workers in corridor (W_{exp})	138	282
Total exposed population ($Pop_A + W_{\text{exp}}$)	187	331
PLL (expected fatalities)	109	189

the upper tail. VaR95 ranges from about 0.1 expected fatalities at Goldex to roughly 150 at Chirano, while P99 ranges from ≈ 0.15 to nearly 200 expected fatalities. The league-table plot of P90–P99 intervals with VaR95 markers makes this contrast clear: a small group of sites concentrates at the upper end of the axis, whereas most sites cluster near zero with very short high-percentile ranges. In terms of upper-tail behaviour, the portfolio is therefore highly heterogeneous, even though all sites were evaluated using the same release-model procedure and the same corridor-parameter grid.

To structure the discussion, we group the 17 dams into high-, moderate- and low-consequence sites using combined tail diagnostics rather than a single threshold metric. VaR95 is used as the main ranking metric, while the P90–P99 range, ES95, scenario-exceedance-curve shape and boxplot behaviour are used to identify sites with long-tail behaviour. This combined interpretation is important because some sites may have low typical or VaR95-level consequences but still exhibit high-consequence outcomes for a small subset of parameter combinations in the upper tail. These labels are study-specific and are not intended to replace the formal consequence classification categories used in standards such as the GISTM [2]. The high-consequence group, conditional on failure, comprises: Loulo-Goukoto, Geita, Edikan, Siguiri and Chirano. For these tailings dams, VaR95 lies between ≈ 71 and 148 expected fatalities, P90 between ≈ 62 and 129, and P99 between ≈ 91 and 197. The P90–P99 spans are wide, indicating that many parameter combinations produce non-negligible loss of life once failure is assumed. The moderate-consequence group consists of KCGM, Kuranakh and Driefontein. KCGM and Kuranakh have VaR95 between roughly 22 and 37 expected fatalities, whereas Driefontein is included in the moderate-consequence group because of its long-tail profile: although its VaR95 remains below one expected fatality, its P99 reaches about 60 and its ES95 is substantially higher, indicating that a small subset of scenarios produces much larger consequences. The remaining nine sites: South Deep, Fosterville, Mponeng, Young-Davidson, Haile, Ergo, Goldex, Aurizona and Yanfolila, form a low-consequence group with VaR95 consistently below ≈ 7 and P99 below ≈ 9 expected fatalities. Appendix A, Table A1 provides an auxiliary descriptive listing of the original disclosure-based consequence or hazard classes reported by Franks et al. [10] alongside the study-specific conditional PLL tiers. Because the disclosed classes originate from different classification systems and may incorporate

environmental, economic, infrastructure or other consequences beyond loss of life, no direct equivalence or agreement is inferred. The table is not used for model validation, tier assignment or prioritisation.

The scenario-exceedance curves are consistent with this three-tier structure. High-consequence sites exhibit relatively shallow scenario-exceedance curves: scenario-exceedance fractions remain non-negligible (much greater than 5%) over tens to more than one hundred expected fatalities before approaching the 5% VaR95 level. Moderate-consequence sites show steeper curves. KCGM and Kuranakh cross the 5% scenario-exceedance level at approximately 20–40 expected fatalities, whereas Driefontein crosses it below one expected fatality but retains a pronounced upper tail. Low-consequence sites have very steep scenario-exceedance curves that approach zero for PLL values of only a few expected fatalities, indicating that only a small fraction of parameter combinations results in any appreciable exposure.

The VaR95-ES95 cross-plots indicate that VaR95 and ES95 are close for most sites at the current scenario-set resolution, implying that the VaR95 threshold is broadly representative of the threshold-defined upper-tail mean. Driefontein is a clear exception: VaR95 remains below one expected fatality while ES95 is much larger, reflecting a distribution in which most parameter combinations yield small PLL but a small subset produces markedly higher consequences. We therefore use VaR95 as the primary tail metric for tiering and ranking, while retaining ES95 to identify sites with heavier upper tails.

Finally, the boxplots of PLL across the scenario set for each site reveal strongly right-skewed distributions whose spread mirrors the consequence grouping. High-consequence sites show large interquartile ranges, high medians (already in the tens of expected fatalities), and numerous extreme outliers. Moderate-consequence sites have medians in the single-digit range and shorter whiskers, while low-consequence sites display medians below about 2 expected fatalities with tight boxes and only occasional outliers. Together, these patterns justify treating the three groups as distinct consequence classes and motivate the more detailed, site-by-site analysis of scenario-exceedance curves and tail behaviour in the following subsections.

3.2. Site-level ranking from PLL percentiles

Fig. 4 summarises the scenario-set outcomes at site level as a “league table”. Each horizontal bar shows the range between the 90th and 99th percentiles of PLL (P90–P99), while the square marker indicates the 95th percentile (P95), which is identical to VaR95 in our notation. Sites are ordered vertically by VaR95, with high-consequence sites at the top, followed by moderate-consequence and then low-consequence sites.

Fig. 4 shows PLL percentiles and highlights a compact group of five high-consequence sites whose upper percentiles lie well above those of the rest of the portfolio. Chirano ranks first, with $P90 \approx 129$, $VaR95 \approx 148$ and $P99 \approx 186$ expected fatalities. Siguiri follows with $P90 \approx 101$, $VaR95 \approx 132$ and $P99 \approx 197$, while Edikan, Loulo-Gouankoto and Geita form a second tier with VaR95 values of about 94, 83 and 71 expected fatalities, respectively, and P99 values between roughly 91 and 140. These sites have long P90–P99 bars that span tens of expected fatalities, indicating that many combinations of release fraction and corridor parameters produce severe loss-of-life outcomes once failure is assumed.

Moderate-consequence sites occupy an intermediate region of the league table. KCGM has $VaR95 \approx 37$ expected fatalities, with $P90 \approx 10$ and $P99 \approx 46$, while Kuranakh has $VaR95 \approx 22$, $P90 \approx 17$ and $P99 \approx 35$. Driefontein is more unusual: its VaR95 is below one expected fatality (≈ 0.73) and $P90 \approx 0.60$, yet its P99 reaches roughly 60 expected fatalities. In the plot this appears as a long, right-extending bar with a marker close to the origin, reflecting a distribution in which most parameter combinations yield very small PLL but a small subset produces much larger consequences. This behaviour is consistent with a corridor geometry in which only the largest releases and longest runout lengths intersect substantial population.

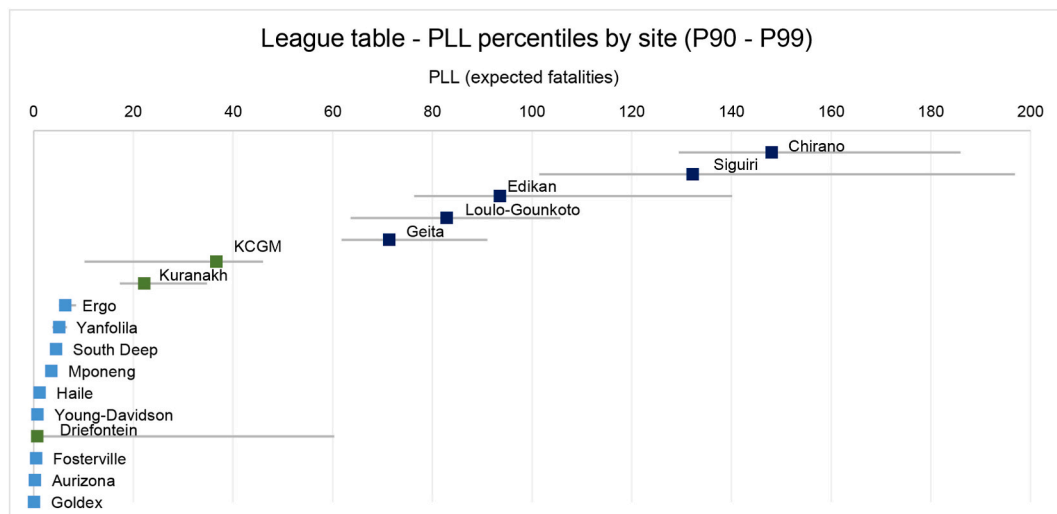


Fig. 4. League-table plot of PLL percentiles across the scenario set by site (P90–P99) with VaR95 markers.

Low-consequence sites cluster near the origin with short P90-P99 intervals. All have VaR95 below about 6.4 expected fatalities, and P99 never exceeds ≈ 8.5 . At the lowest end of the spectrum, Goldex, Aurizona and Fosterville have P99 values below one expected fatality (≈ 0.15 , 0.34 and 0.67 , respectively), and their P90 and VaR95 values are an order of magnitude smaller still. The narrow bars for these dams suggest that even the most pessimistic combinations within our scenario set generate only limited loss-of-life exposure, and that this conclusion is robust to modelling assumptions about release fraction and corridor geometry.

Interpreted as a portfolio screening tool, Fig. 4 therefore serves two functions. First, the position of the VaR95 markers provides a transparent ranking of sites by upper-tail consequence, clearly distinguishing the small number of tailings dams where conditional fatalities could plausibly reach the hundreds from the majority where even P99 upper-tail outcomes remain in the single digits. Second, the width of the P90-P99 intervals offers a simple visual proxy for parameter sensitivity: wide bars (e.g., Chirano, Siguiri, Driefontein) flag sites where PLL consequence estimates are highly sensitive to uncertain model choices, whereas narrow bars (several of the low-consequence sites) indicate robustly low PLL across the scenario set. In the following subsections, we complement this snapshot with the full scenario-exceedance curves, which reveal how these upper-tier metrics arise from the underlying distributions.

3.3. Scenario-exceedance behaviour and VaR95 across consequence groups

The scenario-exceedance curves in Fig. 5 provide a complementary view of the PLL outcomes across the deterministic parameter set. For a PLL threshold q , the scenario-exceedance fraction is the proportion of the 144 parameter combinations for which $PLL > q$. Each curve is therefore a stepwise, monotonically decreasing function of PLL. The horizontal dashed line at 5% identifies the scenario-set VaR95 level, meaning that approximately 5% of the enumerated combinations exceed this value; it does not represent a 5% physical probability of occurrence.

Fig. 5(a) shows scenario-exceedance curves for the five high-consequence sites. Chirano and Siguiri clearly occupy the upper envelope: their curves remain above 10% up to PLL values of roughly 140–150 expected fatalities and cross the 5% line at VaR95 ≈ 148

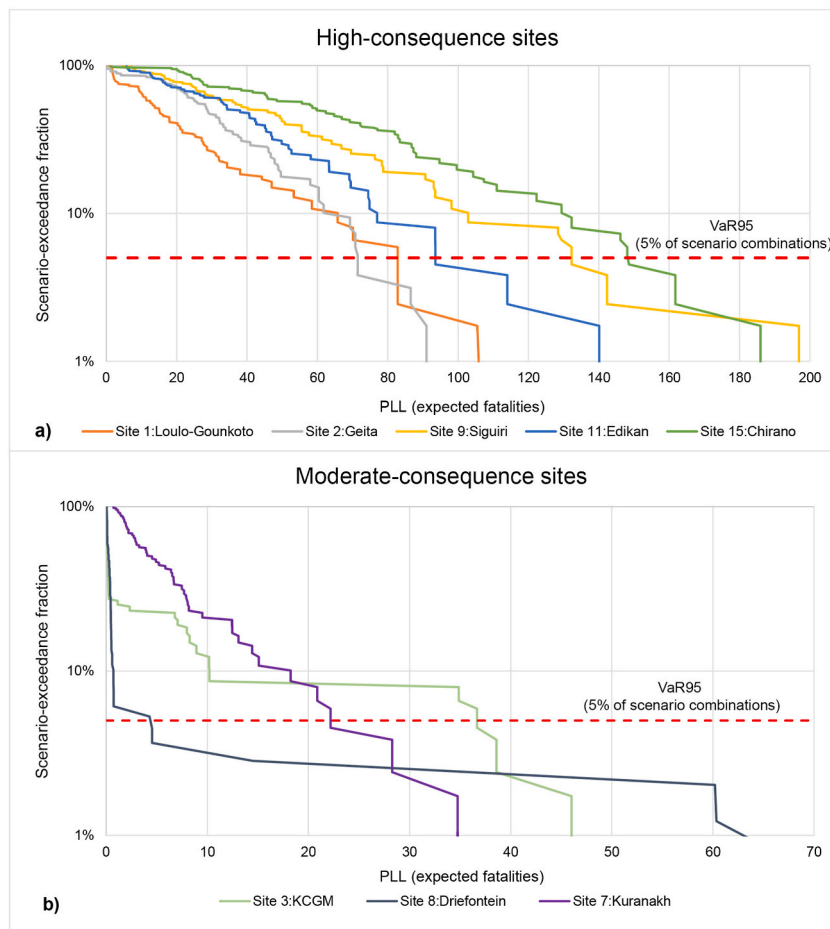


Fig. 5. Scenario-exceedance curves of PLL (expected fatalities) across the deterministic parameter set by consequence group: (a) high-consequence sites; and (b) moderate-consequence sites. The horizontal dashed line identifies scenario-set VaR95, exceeded by approximately 5% of the enumerated parameter combinations.

and ≈ 132 expected fatalities, respectively. Edikan and Loulo-Goukoto form an intermediate tier, with $\text{VaR}_{95} \approx 94$ and ≈ 83 expected fatalities and scenario-exceedance fractions still well above 10% for PLL levels around 70–80. Geita has the lowest curve within the high-consequence group, with $\text{VaR}_{95} \approx 71$ and exceedances already below 10% by ≈ 60 expected fatalities. All five curves decline gradually rather than dropping sharply, indicating that a wide range of parameter combinations can produce moderate to high PLL once failure is assumed. This behaviour is consistent with long, valley-confined runout paths that intersect sizeable downstream populations for many combinations of release fraction and corridor geometry.

Moderate-consequence sites (Fig. 5(b)) display noticeably steeper scenario-exceedance curves and lower VaR_{95} values, but still non-negligible conditional PLL. KCGM has the highest tail within this group: scenario-exceedance fractions remain above 10% up to $\text{PLL} \approx 20$, and the curve crosses the 5% line at $\text{VaR}_{95} \approx 37$ expected fatalities. Kuranakh has $\text{VaR}_{95} \approx 22$, with a comparatively steep decline. Scenario-exceedance fractions fall below 10% by ≈ 18 expected fatalities and approach 1% by ≈ 35 . Driefontein behaves differently: its curve is very steep near the origin, with VaR_{95} only ≈ 0.73 and scenario-exceedance fractions below 10% already at $\text{PLL} \approx 1$. However, a long tail formed by a small share of the parameter combinations extends to $\text{PLL} 60$ (P99), indicating that some combinations generate much larger consequences than the bulk of the scenario set. This pattern suggests a corridor where only extreme releases and long runout lengths reach populated areas.

Low-consequence sites (Appendix A, Figure A1) have scenario-exceedance curves clustered close to the origin, with VaR_{95} values below ≈ 6.4 expected fatalities for all nine tailings dams. For Goldex, Aurizona and Fosterville, VaR_{95} is below one expected fatality (≈ 0.11 , 0.24 and 0.52 , respectively), and scenario-exceedance fractions drop below 5% at PLL levels that remain well under a single expected fatality. South Deep, Mponeng, Yanfolila and Ergo exhibit somewhat longer tails, but even for these sites the 5% scenario-set exceedance level lies between ≈ 3.6 and ≈ 6.3 expected fatalities, and the curves approach the 1% line by $\text{PLL} \approx 5$ –8. Overall, the steepness and limited horizontal extent of the low-consequence curves indicate that only a narrow subset of parameter combinations yields any material exposure, and that even pessimistic scenarios remain in the single-digit range of expected fatalities.

Viewed together, the scenario-exceedance curves clarify how the league-table statistics arise from the underlying distributions.

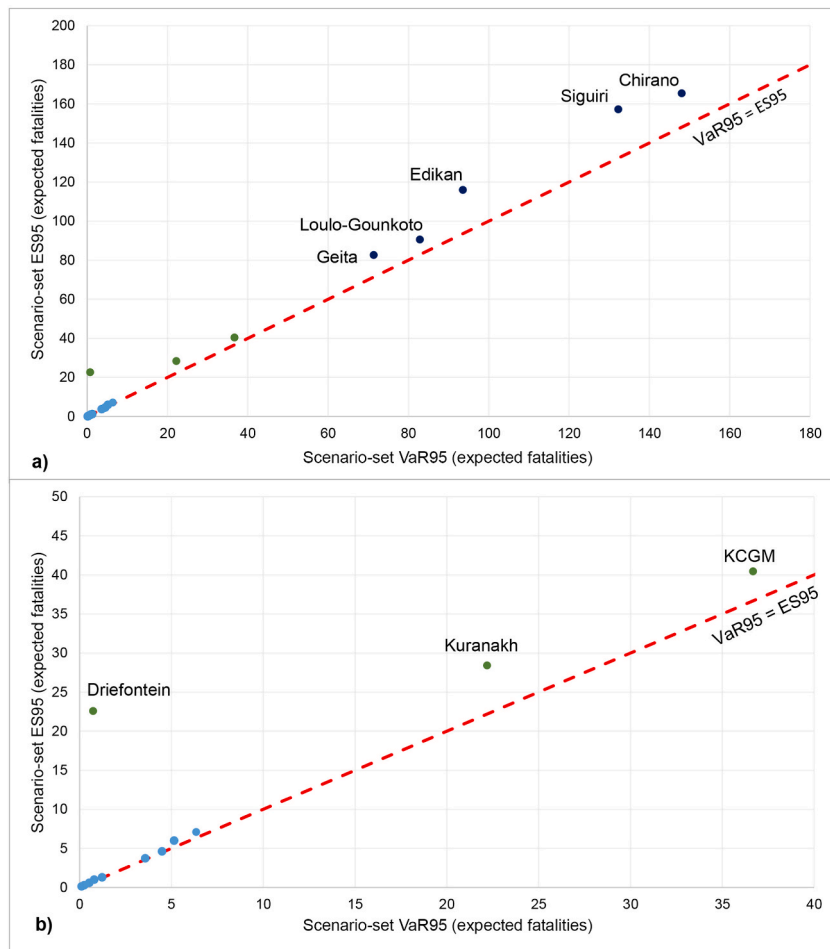


Fig. 6. Relationship between VaR_{95} and ES_{95} for PLL (expected fatalities) across the scenario set: (a) high- and low-consequence sites; (b) moderate- and low-consequence sites. The red dashed line shows the 1:1 relationship ($\text{VaR}_{95} = \text{ES}_{95}$).

Gradual, slowly decaying curves in the high-consequence group reflect corridors where many combinations of release fraction and thickness parameters intersect dense downstream populations, leading to a broad range of PLL values with substantial share of scenario combinations at high consequences. In contrast, steep curves in the moderate- and low-consequence groups indicate that results are dominated by a small set of severe parameter combinations, with most runs producing little or no exposure because the modelled runoff either stops before reaching major settlements or traverses sparsely populated terrain. These patterns reinforce the three consequence classes introduced in Sections 3.1-3.2 and show that the ranking by VaR95 is supported not only by point estimates but also by the full shape of the scenario-set distributions.

3.4. Tail severity: relationship between VaR95 and ES95

Fig. 6 compares VaR95 with ES95, defined here as the threshold-based upper-tail mean of PLL outcomes at or above the VaR95 level. Fig. 6(a) shows the high- and low-consequence sites, and Fig. 6(b) shows the moderate- and low-consequence sites, with the low-consequence cluster repeated in both panels to keep the axes comparable. The red dashed line represents the 1:1 relationship $\text{VaR95} = \text{ES95}$. Points located close to this line have tails in which the threshold-defined upper-tail mean is similar to the VaR95 threshold. Points above the line indicate heavier tails, where the threshold-defined upper-tail mean is higher than the VaR95 threshold itself.

Across the portfolio, most sites lie very close to the $\text{VaR95} = \text{ES95}$ line, indicating that ES95 is numerically similar to VaR95 at the scenario-set resolution used here. High-consequence sites cluster in the upper right of Fig. 6(a). Chirano and Siguiri plot slightly above the 1:1 line, indicating that for these tailings dams the threshold-defined upper-tail mean is modestly higher than the VaR95 threshold, consistent with the gradual scenario-exceedance curves and long P90-P99 ranges described in Sections 3.2-3.3. Edikan, Loulo-Goukoto and Geita are even closer to the line, with ES95 only marginally exceeding VaR95. Thus, although these sites have large conditional PLL, their tails are not extremely heavy in the sense that the very worst runs are only somewhat more severe than the 95th-percentile scenario.

Moderate-consequence sites (Fig. 6(b)) show a similar pattern at lower absolute consequence levels. KCGM and Kuranakh lie

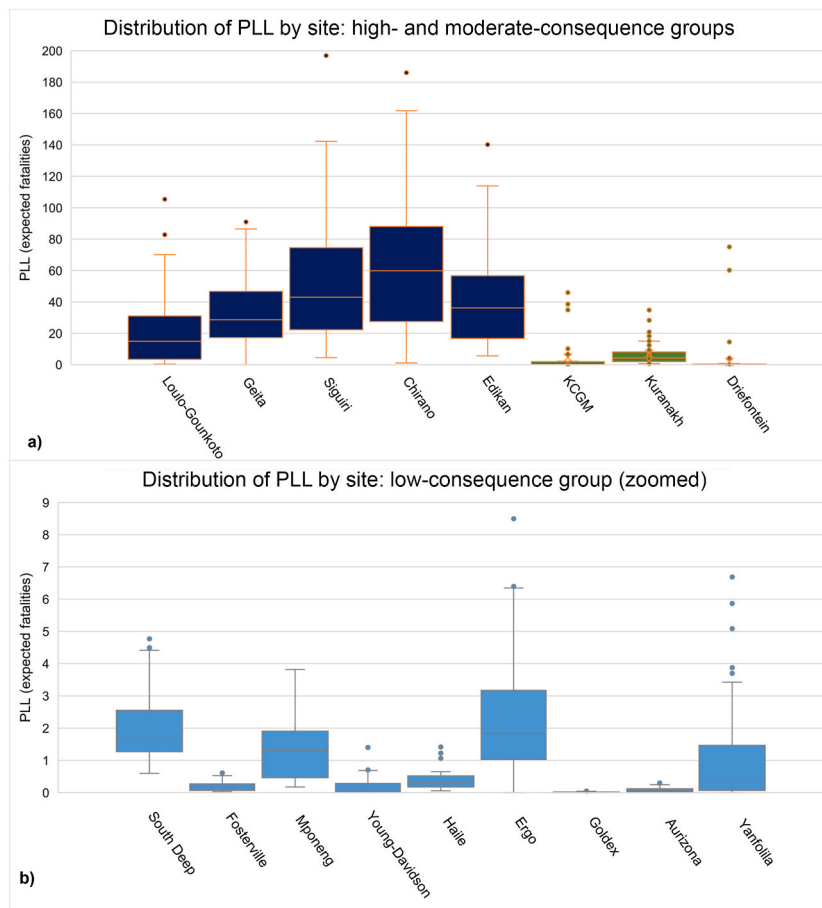


Fig. 7. Distributions of PLL (expected fatalities) across the scenario set, by site: (a) high- and moderate-consequence groups; (b) low-consequence group (zoomed).

slightly above the $\text{VaR95} = \text{ES95}$ line, reflecting the modest tail extension seen in their scenario-exceedance curves. Driefontein is a clear exception: VaR95 remains below one expected fatality, while ES95 is much larger because the threshold-defined upper-tail set includes a small number of severe runs. This indicates a thin-bulk, long-tail distribution, where most parameter combinations produce very low PLL but a limited subset reaches substantially higher exposed populations. Low-consequence sites cluster near the origin, with both VaR95 and ES95 in the single-fatality range or below, and they sit essentially on the 1:1 line, indicating thin tails in which the threshold-defined upper-tail mean is only slightly higher than VaR95 .

From a portfolio-management perspective, these scatterplots (Fig. 6(a) and (b)) show that in this particular scenario set the choice between VaR95 and ES95 makes little difference to the relative ranking of most sites: tailings dams that have large VaR95 also have proportionally large ES95 , and most points remain close to the $\text{VaR95} = \text{ES95}$ line. VaR95 is therefore an adequate primary ranking metric for portfolio-level screening in our case. ES95 mainly acts as a diagnostic check on tail behaviour, especially for sites such as Driefontein, where the small number of severe runs is not fully represented by VaR95 alone. More generally, points located substantially above the 1:1 line would indicate genuinely heavy tails, where the very worst scenarios are much more severe than the 95th-percentile case.

3.5. Within-site variability: full PLL distributions

Fig. 7 summarises the distribution of PLL (expected fatalities) across the full scenario set for each site, using boxplots. For every dam, the box spans the interquartile range (IQR, P25–P75), the horizontal line marks the median, the whiskers extend to the most extreme values within $1.5 \times \text{IQR}$ of the quartiles, and points beyond the whiskers denote outlying runs. All values remain conditional on dam failure and on the set of parameter scenarios.

For the high-consequence sites (Fig. 7(a)), the boxplots show that large consequences are not confined to a handful of extreme runs but occur across a substantial part of the scenario set. Chirano and Siguiri have the highest medians, at about 60 and 43 expected fatalities, respectively, with very wide IQRs ($\approx 27\text{--}88$ for Chirano and $\approx 23\text{--}71$ for Siguiri) and P90 values around 130 and 100. Their upper whiskers and outliers extend to nearly 190–200 expected fatalities, confirming the broad right-skewed distributions and elevated VaR95 and ES95 values seen in the scenario-exceedance curves. Edikan and Geita form an intermediate tier: their medians (≈ 36 and ≈ 29) and IQRs ($\approx 17\text{--}54$ and $\approx 18\text{--}47$) still imply double-digit PLL for a large share of runs, with P90 around 76 and 62 and maxima near 140 and 91. Loulo-Goukoto has a lower median of ≈ 15 expected fatalities, but its long upper whisker (P90 ≈ 64 , maximum ≈ 106) indicates that a sizeable minority of parameter combinations still produces consequences comparable to those at Geita.

Moderate-consequence sites, plotted on the same scale in Fig. 7(a), show a different structure. Kuranakh has a median PLL of about 4.3 expected fatalities, with an IQR of $\approx 2.1\text{--}8.1$ and P90 ≈ 17 , placing it just below the high-consequence sites in terms of typical outcomes but clearly separated from the low-consequence group. KCGM and Driefontein, by contrast, have medians close to zero (≈ 0.14 and ≈ 0.15) and relatively compact boxes and P90 values (≈ 10 for KCGM and ≈ 0.6 for Driefontein), yet both exhibit very long upper whiskers and isolated outliers. In particular, Driefontein reaches a maximum of ≈ 75 expected fatalities despite its very low P90, while KCGM has outliers up to ≈ 46 . This pattern, small medians and P90 but occasional large outliers, is consistent with the steep scenario-exceedance curves and the $\text{VaR95}/\text{ES95}$ behaviour for these sites, where most runs yield little or no exposure but a small subset of high-release, long-runout combinations generates disproportionately severe consequences.

The zoomed panel for low-consequence sites (Fig. 7(b)) confirms that their PLL distributions across the scenario set are tightly concentrated near zero. Medians range from ≈ 0.005 expected fatalities at Goldex and ≈ 0.05 at Aurizona to $\approx 1.3\text{--}1.9$ at South Deep, Mponeng and Ergo. P90 values remain below about 5 expected fatalities for all sites (≈ 5.1 at Ergo, ≈ 4.4 at South Deep and ≈ 3.7 at Yanfolila), and maxima are in the single-digit range except for a few isolated points. Fostererville, Young-Davidson, Haile, Goldex and Aurizona show particularly compact boxes with P90 well below one expected fatality, indicating that even pessimistic parameter combinations produce only marginal conditional PLL. South Deep, Mponeng, Ergo and Yanfolila have somewhat broader IQRs and longer whiskers, but still fall well below the high- and moderate-consequence sites in both median and upper-tail values.

Taken together, the boxplots reinforce the patterns seen in the league table, scenario-exceedance curves and $\text{VaR95}\text{--}\text{ES95}$ scatterplots. Sites with high upper-tail severity and strongly skewed distributions (notably Chirano, Siguiri and, to a lesser extent, Edikan, Geita, Kuranakh and Driefontein) are characterised by skewed distributions with long upper whiskers and pronounced outliers. In contrast, low-consequence sites have compact, near-symmetric boxes clustered close to zero, reflecting both low typical PLL and limited sensitivity to parameter choices. In modelling terms, large within-site variability arises where combinations of high release, long decay length and corridor width generate greater effective depth and/or longer reach across populated areas than low-release or short-runout combinations. Where modelled runout paths are short or traverse sparsely populated terrain, the PLL distributions across the scenario set remain sharply bounded, and both median and tail metrics stay in the single-fatality range.

3.5.1. Sensitivity of high-consequence PLL to warning-response assumptions

Table 7 presents the warning-response sensitivity results for the five high-consequence sites. Relative to the reference depth–fatality curve, warning issued at the time of failure reduced VaR95 and ES95 by 22.4–34.4%, while a 5 min advance warning reduced both metrics by 43.9–59.2%. The magnitude of this variation confirms that human-response assumptions are an important source of uncertainty in the absolute PLL estimates.

Under the warning-at-failure variant, VaR95 remained between approximately 47 and 115 expected fatalities across the five sites. Under the 5 min warning variant, the corresponding range was approximately 29–83 expected fatalities. Chirano and Siguiri remained the two highest-consequence sites in all variants, followed by Edikan, Loulo-Goukoto and Geita. The rank order remained unchanged because the same sensitivity multipliers were applied to each run and site.

These findings should not be interpreted as forecasts of the benefits of warning systems at the individual facilities. The transferred multipliers represent the range produced by alternative evacuation assumptions for the Brumadinho case. Actual effectiveness would depend on mudflow arrival time, warning reliability, evacuation-route availability, population awareness, building characteristics and the spatial distribution of people when failure occurs. The analysis nevertheless demonstrates that the reference PLL values may vary materially under plausible alternative response assumptions, even where the portfolio ranking remains stable.

3.5.2. Sensitivity to release-corridor-width dependence

The equal-weight full-factorial baseline includes every combination of release fraction and corridor half-width but does not assign higher likelihood to any dependence structure. Two bounding rank-coupling tests were therefore used to examine the effect of possible release-width dependence. Positive coupling, in which progressively larger releases were paired with progressively wider corridors, increased VaR95 at the five high-consequence sites by approximately 25.6–49.8% relative to the baseline. The opposite coupling reduced their VaR95 values by approximately 31.6–55.6%. ES95 changed by approximately +10.2% to +25.3% under positive coupling and by approximately –40.8% to –51.4% under the opposite bound.

Despite these changes in absolute magnitude, the portfolio ranking was comparatively stable. Spearman rank correlations between the baseline and dependence-test VaR95 values were $\rho = 0.983$ for positive coupling and $\rho = 0.988$ for the opposite coupling. Corresponding ES95 rank correlations were $\rho = 0.998$ and $\rho = 0.953$. Chirano, Siguiri, Edikan, Loulo-Gouunkoto and Geita remained the five high-consequence sites in both tests, although their ordering within the group changed slightly.

These results show that assumed release-width dependence can materially affect absolute VaR95 and ES95 estimates, but the principal portfolio grouping is less sensitive. The stress tests do not identify the actual dependence structure at any facility. Rather, they bound its potential influence and reinforce the interpretation of the reported metrics as design-based screening summaries rather than calibrated probabilistic quantities.

3.6. Tail PLL severity versus downstream population and stored tailings volume

Fig. 8 examines how cross-site differences in tail PLL severity, represented by ES95, relate to estimated stored tailings volume and

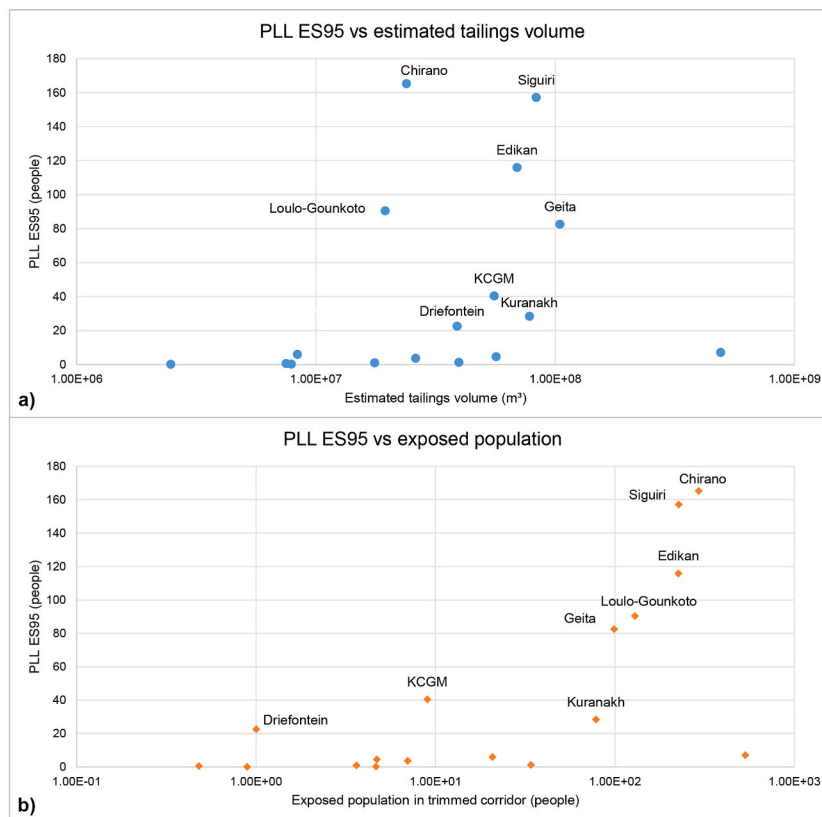


Fig. 8. Relationship between tail PLL severity (ES95; expected fatalities) and (a) estimated stored tailings volume and (b) standardised reference-run exposed population, as defined in Section 2.6. The x-axes are shown on a log₁₀ scale to accommodate orders-of-magnitude variation. Pairwise and partial Spearman correlations are reported in Table 8. Point labels identify the high- and moderate-consequence sites; unlabelled points correspond to low-consequence sites.

the standardised reference-run exposed population defined in Section 2.6. The pairwise Spearman association between ES95 and stored volume was moderate ($\rho = 0.60$), whereas the association between ES95 and exposed population was stronger ($\rho = 0.78$). Exposed population and stored volume were also positively associated ($\rho = 0.67$), indicating that comparison of the two bivariate coefficients alone does not isolate their separate relationships with ES95.

Partial Spearman analysis produced a clearer contrast. After controlling for stored volume, ES95 retained a moderate-to-strong association with exposed population ($\rho_{\text{partial}} = 0.64$; permutation $p = 0.006$). After controlling for exposed population, the remaining association between ES95 and stored volume was weak ($\rho_{\text{partial}} = 0.16$; permutation $p = 0.529$). The leave-one-site-out partial correlation ranged from 0.48 to 0.79 for exposed population controlling for volume and from 0.04 to 0.33 for stored volume controlling for population. The contrast therefore did not depend solely on one facility or on the largest-volume observation.

These results indicate that downstream population provides more independent rank information about tail PLL severity than stored volume within this 17-site portfolio. They strengthen the case for exposure-aware screening and show why volume alone is insufficient for prioritisation. Nevertheless, the small sample and screening-derived variables preclude a universal or causal interpretation. The findings should therefore be read as portfolio-specific evidence of a stronger exposure association rather than proof that exposure dominates consequence magnitude in every TSF setting.

The pairwise correlations, partial correlations, permutation-test results and leave-one-site-out ranges are summarised in Table 8.

3.7. Synthesis, comparison with literature and policy implications

Taken together, the percentile league table, scenario-exceedance curves, VaR95-ES95 plots and boxplots identify five robustly high-consequence sites (Chirano, Siguiri, Edikan, Loulo-Goukoto and Geita), three moderate-consequence sites (KCGM, Kuranakh and Driefontein) and nine low-consequence sites. The high-consequence group has VaR95 values of 71–148 expected fatalities, wide P90–P99 intervals and strongly skewed distributions. KCGM and Kuranakh have intermediate VaR95 values, whereas Driefontein shows a thin-bulk, long-tail pattern. The remaining sites have medians below two expected fatalities and VaR95 below 7. The three-tier grouping is supported consistently across all four diagnostics, indicating that it is not driven by a single metric. Exposed population also retained a stronger partial association with ES95 ($\rho = 0.64$) than stored volume ($\rho = 0.16$), although the 17-site sample limits generalisation.

The upper-tail conditional PLL values at the high-consequence sites are of the same order of magnitude as fatality totals reported for some major historical tailings failures, although they are scenario-based expected values rather than observed event outcomes [1,4,19]. The three-tier pattern is also qualitatively consistent with the GISTM, which distinguishes failures capable of causing mass casualties from those with more localised impacts [2]. The study-specific tiers do not replace formal classifications, but they separate facilities with plausible community-scale consequences from those remaining in the single-digit range given failure.

Leave-one-out cross-validation of the 54-event release-volume dataset produced a cross-validated R^2 of 0.607, compared with an in-sample R^2 of 0.646. RMSE and MAE were 0.583 and 0.452 $\log_{10}(\text{m}^3)$, equivalent to multiplicative factors of approximately 3.83 and 2.83, while mean prediction error was 0.005 $\log_{10}(\text{m}^3)$, indicating little directional bias. Empirical coverage was 75.9% for the nominal 16th–84th percentile interval and 92.6% for the nominal 2.5th–97.5th percentile interval. When Brumadinho was withheld, its median predicted release was $3.65 \times 10^6 \text{ m}^3$ versus an observed $9.57 \times 10^6 \text{ m}^3$, which remained within the fold-specific 16th–84th percentile interval of approximately $0.98\text{--}13.63 \times 10^6 \text{ m}^3$. These results show that the regression captures the broad scale of release volumes but retains substantial event-level uncertainty, supporting its use as a scenario generator rather than a precise deterministic predictor. Table 9 reports the complete metrics.

The Brumadinho failure was also used as a screening-level plausibility check of routing and workforce exposure (Fig. 9; Table 10). Released volume was fixed at $9.7 \times 10^6 \text{ m}^3$, consistent with Robertson et al. [8] and Lumbroso et al. [7], and the same workflow used for the 17-site portfolio produced an applied corridor length of 9.028 km. Because this length is the shorter of the theoretical stopping distance and the available digitised centerline, it represents the applied screening-corridor extent rather than an independently validated maximum runout. Published accounts indicate an approximately 10 km downstream path to the Paraopeba River [7,9], making the modelled extent about 0.97 km, or 9.7%, shorter.

The comparison with Lumbroso et al. [7] also clarifies the difference in model scope. Their reconstruction combined an



Fig. 9. Brumadinho (2019) tailings dam failure: (a) pre-failure setting, (b) post-failure inundation footprint, and (c) modelled hazard corridor (Scenario B1).

EMBREA-MUD breach hydrograph with a two-dimensional MIKE 21 Bingham-flow model and an approximately 12 m terrain model to simulate time-varying depth, velocity and arrival time. The present framework instead uses a fixed released volume, a MERIT-guided centerline, a constant-width envelope and a longitudinal effective-thickness function. The similarity in downstream extent supports first-order route and runout plausibility for this case, but does not validate local width, depth, velocity, arrival time or deposition.

Residential population grids inadequately represent industrial occupancy [41], so an explicit workforce layer was added. Published loss-of-life modelling assumed approximately 400–430 workers on site at the time of failure [7]. Accordingly, the workforce input was set to $N_{\text{work}} = 400$ and allocated within a mine working-area polygon. To bound uncertainty in workforce location, Scenario B₁ distributed workers across a relatively wide polygon, whereas Scenario B₂ concentrated them within a smaller polygon (Fig. 9; Table 10). WorldPop-only exposure was low in both scenarios ($Pop_A = 49$). The corridor intersected $W_{\text{exp}} = 138$ workers in Scenario B₁ and $W_{\text{exp}} = 282$ in Scenario B₂, producing total exposed populations ($Pop_A + W_{\text{exp}}$) of 187 and 331 and corresponding PLL values of 109 and 189 expected fatalities. The comparison therefore supports two screening-level observations: the applied downstream extent is of the same order as the published approximately 10 km reconstruction, and conditional PLL is highly sensitive to workforce location. It remains a first-order plausibility check rather than validation of local flow width, depth, velocity, arrival time, deposition or fatality outcomes.

The scenario-set framework complements single-scenario breach modelling by making parameter sensitivity visible and providing a common basis for portfolio comparison. Wide P90–P99 ranges, slowly decaying scenario-exceedance curves and long boxplot tails identify sites where release and corridor assumptions strongly affect PLL. VaR95 provides the primary ranking metric, while ES95 identifies heavier tails such as Driefontein. These metrics support prioritisation but cannot determine capital allocation, regulatory compliance or risk acceptance without site-specific assessment of failure likelihood, breach behaviour, downstream exposure and emergency response [2,15,39].

A conservative Stage 2 escalation rule is therefore proposed. A facility is flagged for Stage 2 detailed assessment when.

- scenario-set VaR95 is at least 10 expected fatalities;
- VaR95 is below 10 but ES95 or P99 is at least 10, indicating a long-tail case;
- results are materially sensitive to warning, workforce exposure, routing or parameter dependence; or
- data-quality limitations or an existing jurisdictional classification independently justify further investigation.

The threshold of 10 expected fatalities is a study-specific screening anchor that broadly corresponds to the GISTM transition from High to Very High on the life-loss dimension and, with the long-tail override, separates the low-consequence sites from the moderate- and high-consequence cases [2]. It is neither a regulatory threshold nor a tolerable-risk criterion. Because the framework estimates consequences conditional on failure rather than annual failure probability, its PLL metrics cannot be compared directly with annual risk-tolerability criteria. Stage 2 should include a site-specific breach hydrograph, two-dimensional non-Newtonian runout modelling, refined topography, transient and workforce exposure, warning and evacuation analysis, and a formal risk assessment combining consequences with failure likelihood and applicable requirements to reduce risk as low as reasonably practicable [2,15].

GISTM requires operators to conduct breach analyses for credible failure modes and assess downstream exposure and vulnerability (Requirements 2.3 and 2.4). Under Requirement 4.1 and Annex 2, operators select the highest applicable consequence class across population at risk, potential loss of life, environmental, health, social, cultural, infrastructure and economic effects. Principle 15 and Requirement 15.3 link disclosure to public accountability and call for standardised, independent, industry-wide and publicly accessible information repositories [2].

Hopkins and Kemp [12] regard GISTM as an important advance while describing compromises involving industry influence, accountability, affected-community rights, disclosure and independent assurance. GISTM provides Independent Tailings Review Boards and Dam Safety Reviews, yet they question whether company-arranged assurance and summary disclosure enable sufficient external review. Credibility therefore depends on whether external actors can examine the evidence and assumptions behind operator-generated assessments.

Within this governance context, the proposed open-data framework provides an independently reproducible Stage 1 screen. Common datasets, explicit assumptions and archived scenarios allow external actors to compare facilities, test rankings and question operator-generated classifications. The framework cannot establish a formal GISTM classification, verify conformance or replace site-specific breach analysis, independent technical review, affected-community engagement, emergency planning or regulatory judgement.

PLL covers only fatality consequences. It does not quantify displacement, livelihood disruption, health effects, cultural heritage loss, environmental damage, infrastructure loss or economic disruption. Because GISTM assigns the highest applicable consequence class across all dimensions, low screening-level PLL does not imply low overall consequence [2,12]. The framework cannot substitute for social, environmental or local economic impact assessments.

Several limitations affect the absolute PLL estimates. First, the same depth-fatality curve is applied to all sites. Warning-response variants reduced VaR95 and ES95 by approximately 22–34% for warning at failure and 44–59% for a 5 min advance warning, while preserving the ranking because common multipliers were applied. These transfer-based bounds do not represent site-specific building stock, shelter, warning reliability, evacuation capacity, arrival time or temporal occupancy. The reference curve may therefore overestimate PLL where warning, evacuation or shelter are effective and underestimate it where warning is absent, occupancy is concentrated or vulnerable buildings predominate [7,16,25].

Second, routing uses a MERIT-guided one-dimensional centerline, a constant-width envelope and an exponential effective-

thickness profile rather than a two-dimensional hydrodynamic model [7,20,32]. It does not simulate local widening and constriction, bends, lateral spreading, deposition, rheological transitions, velocity or arrival time. Exposure may be overestimated where actual flow is more confined and underestimated where spreading or runout exceeds the selected corridor. The modelled thickness is therefore an exposure proxy, not a locally validated depth prediction.

Third, the release regression is conditional on documented failures, which may overrepresent large or well-investigated events. Leave-one-out validation quantifies predictive performance within the 54-event sample but cannot correct missing-event bias. If smaller failures are systematically underrepresented, the fitted relationship may overstate release severity for some facilities.

Fourth, equal design weights provide balanced coverage of the selected parameter space but do not represent physical likelihood. The full-factorial design also does not establish that the parameters are physically independent. Probabilistic, Bayesian or expert-weighted approaches require defensible marginal distributions and dependence structures that are unavailable consistently across the portfolio [37]. The release-width stress tests materially changed absolute VaR95 and ES95 but preserved strong cross-site rank agreement and the same five high-consequence sites. The reported metrics should therefore be interpreted as design-based sensitivity summaries rather than calibrated probabilities. Archived run-level results can be reweighted if site-specific distributions, expert weights or dependence models become available.

Finally, the cross-site analysis contains only 17 facilities. Partial rank correlations and leave-one-site-out checks reduce confounding and sensitivity to individual observations, but do not establish causation or guarantee transferability. Exposure is also based on one WorldPop realisation at 100 m resolution without explicit treatment of population change or uncertainty [23,28]. Rankings therefore remain conditional on the selected parameter ranges, equal weights and population representation.

These limitations can bias results in either direction. Reporting bias may overstate release severity, omitted transient or workforce populations may understate exposure, and corridor or fatality assumptions may over- or underestimate PLL depending on confinement, spreading, warning and vulnerability. DEM, centerline and population-grid uncertainties may also act in either direction. The values should therefore be treated as assumption-dependent screening estimates rather than consistently conservative predictions. The sensitivity analyses indicate that portfolio grouping is more stable than absolute PLL [7,23,41].

The framework is therefore a Stage 1 portfolio-screening tool. It identifies facilities warranting Stage 2 breach, runout, exposure and evacuation assessment, but does not determine risk acceptability, regulatory compliance or investment decisions. Data-quality limitations, sensitivity findings and governance conditions apply regardless of the numerical result.

4. Conclusion

This study developed an end-to-end, screening-level framework for portfolio-wide consequence assessment of tailings dam failure using Google Earth Engine (GEE) and an Excel regression release model. The workflow estimates dam geometry by DEM cut-and-fill, derives cap-aware release scenarios, routes a simplified MERIT-guided corridor with exponential thickness decay, and computes conditional Potential Loss of Life (PLL) on the WorldPop 100 m grid. It supports portfolio prioritisation given failure rather than breach-hydraulics simulation or replacement of site-specific dam-break and evacuation analyses.

Conditional consequences varied substantially across the 17-site portfolio despite a consistent scenario design of 144 runs per site, with upper-tail PLL spanning nearly three orders of magnitude (VaR95 ~ 0.1 at Goldex to ~ 148 at Chirano; P99 ~ 0.15 to ~ 200). Chirano, Sigui, Edikan, Loulo-Goukoto and Geita formed the high-consequence group (VaR95 ~ 71 – 148 ; broad P90–P99), whereas nine sites remained consistently low (VaR95 $< \sim 7$; P99 $< \sim 9$). Scenario-exceedance curves distinguished broadly severe distributions from cases where severe outcomes occurred only for a small share of parameter combinations. Driefontein illustrates this long-tail profile (P90 < 1 ; P99 ~ 60). VaR95 and ES95 were closely aligned for most sites, supporting VaR95 as the primary screening metric and ES95 as a tail diagnostic. Warning-response variants reduced high-consequence-site VaR95 and ES95 by approximately 22–59%, confirming substantial uncertainty in absolute PLL while preserving the ranking under common transfer-based multipliers. Exposed population retained a stronger association with ES95 after controlling for stored volume ($\rho_{partail} = 0.64$) than stored volume did after controlling for population ($\rho_{partail} = 0.16$). This supports exposure-aware screening and cautions against volume-only prioritisation, although the 17-site sample limits generalisation.

These findings support portfolio governance by providing an independently reproducible first-pass reference for comparing facilities and questioning operator-generated classifications. Explicit inputs, assumptions and documented scenario logic allow operators and external stakeholders, including regulators, investors, civil society organisations and researchers, to reproduce the screen. VaR95, scenario-exceedance-curve shape and sensitivity results identify where detailed breach and two-dimensional runout modelling, emergency planning and data validation are most informative. These outputs support prioritisation and external review but cannot determine GISTM conformance or independently assure facility safety.

Under the proposed escalation rule, facilities with scenario-set VaR95 of at least 10 expected fatalities, or ES95 or P99 of at least 10 despite lower VaR95, are flagged for Stage 2 detailed assessment. Data-quality limitations, sensitivity findings and governance conditions apply regardless of the numerical result. These are consequence-screening triggers, not tolerable-risk, regulatory-acceptance or capital-allocation criteria.

Results remain conditioned on simplified assumptions. Runout is represented by a MERIT-guided centerline corridor with exponential thickness decay, not two-dimensional hydraulics or complex spreading. Released volume comes from regression-based scenarios with residual spread and a cap rather than a time-varying breach hydrograph. Baseline PLL uses a uniform depth-fatality curve on single-year WorldPop 100 m data. Warning-response sensitivity uses transfer-based multipliers for the five high-consequence sites and does not model site-specific warning effectiveness, evacuation capacity, building stock or transient occupancy. DEM-derived

geometry and routing also contain elevation uncertainty and some analyst dependence. The full-factorial scenario set uses equal design weights for balanced, reproducible coverage, but these weights do not represent physical likelihood. Scenario-exceedance curves, VaR95 and ES95 are therefore design-based conditional-consequence summaries rather than estimates of probabilistic annual risk. Bounding release-width tests changed absolute tail metrics but preserved the five-site high-consequence group, indicating that rankings were more stable than numerical PLL values.

These limitations may over- or understate absolute PLL depending on site conditions, reinforcing the framework's role in comparative screening rather than site-specific prediction.

Future work should extend beyond PLL by combining ESA WorldCover and change detection with spatial indicators of displacement, livelihoods, water resources, protected areas, ecological sensitivity, cultural assets and infrastructure. These data would enable multi-impact screening within the existing corridor and scenario-set logic. Until then, PLL remains one component of consequence assessment and cannot substitute for assessment of GISTM's social, environmental and economic dimensions. The framework's policy value lies in making this fatality component transparent and open to external review.

CRedit authorship contribution statement

Vancho Adjiski: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Tomasz Orpiszewski:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **Peter Schwendner:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vancho Adjiski reports a relationship with Zürich University of Applied Sciences, School of Management and Law, Winterthur, Switzerland that includes: non-financial support. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the providers of the publicly available geospatial datasets used in this study, including Copernicus DEM, MERIT Hydro, WorldPop, and ESA WorldCover. The authors also acknowledge Google Earth Engine as the cloud-based geospatial processing platform used to implement the screening workflow.

Appendix A. Additional results and contextual comparison

Figure A1 presents additional scenario-set results generated in this study. Table A1 reproduces the original disclosure-based consequence or hazard classes and classification-system identifiers reported by Franks et al. [10] for descriptive context.

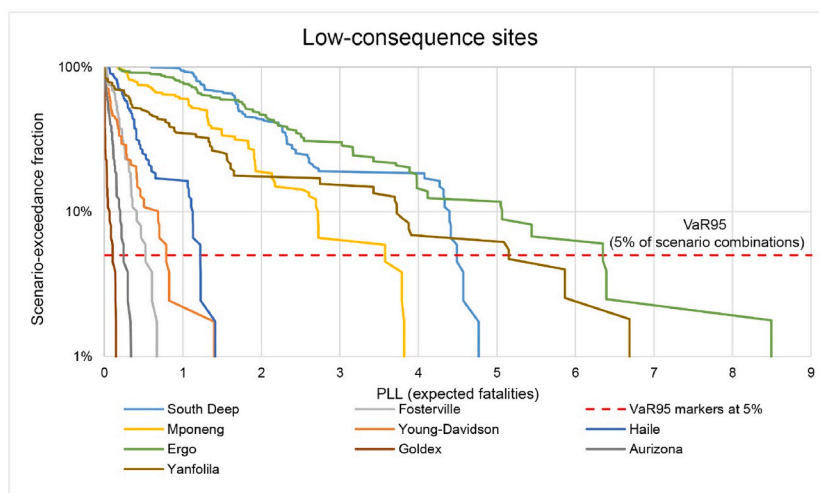


Fig. A.1. Scenario-exceedance curves of PLL (expected fatalities) across the deterministic parameter set for low-consequence sites. The horizontal dashed line identifies scenario-set VaR95, exceeded by approximately 5% of the enumerated parameter combinations.

Table A.1

Auxiliary descriptive listing of this study's conditional PLL tiers and the original disclosure-based consequence or hazard classes reported by Franks et al. [10].

Site	This study: consequence tier (conditional PLL)	Franks: reported class	Franks: classification system
Loulo-Goukoto	High	High	SANS
Geita	High	High	SANS
KCGM	Moderate	High	ANCOLD
South Deep	Low	High B	ANCOLD
Fosterville	Low	Major	AU-VIC
Mponeng	Low	High	SANS
Kuranakh	Moderate	Insignificant	RSA-2
Driefontein	Moderate	High	SANS
Siguiri	High	High	SANS
Young-Davidson	Low	High	CAN-ON
Edikan	High	Hazard Class B/C	GHA
Haile	Low	Low	US-SC
Ergo	Low	High	SANS
Goldex	Low	Medium	CDA
Chirano	High	High	CDA
Aurizona	Low	Very High	CDA
Yanfolila	Low	High	CDA

Note: This study's tier is a study-specific conditional PLL grouping derived from the scenario-set results in Table 6 and the scenario-exceedance, VaR95 and ES95 diagnostics. The Franks et al. [10] column reproduces the original disclosed consequence or hazard class for each facility, together with its reported classification system. The classifications are presented together for descriptive context only. No cross-system harmonisation, equivalence, agreement score or validation inference is made because the underlying schemes differ in terminology, thresholds and consequence dimensions and may incorporate environmental, economic, infrastructure or other impacts in addition to potential loss of life.

Classification-system codes, as reported in the dataset of Franks et al. [10], used in Table A1.

- SANS = South African National Standard 10286 (South African National Standards; consequence classification for mine residue deposits).
- ANCOLD = Australian National Committee on Large Dams (tailings/dam consequence categories under ANCOLD guidance).
- CDA = Canadian Dam Association (dam safety/consequence classification under CDA guidance).
- AU-VIC = Victoria (Australia): Department of Primary Industries, Environmental Guidelines-Management of Tailings Storage Facilities (Nov 2004).
- CAN-ON = Ontario (Canada): Ministry of Natural Resources, Classification and Inflow Design Criteria (2011).
- US-SC = United States (South Carolina): Department of Health and Environmental Control, Dam Safety Program Manual (2012).
- GHA = Ghana: LI 2182-Minerals and Mining (Health, Safety and Technical) Regulations (2012).
- RSA-2 = Russian Federation: State Regulation No. 986 (02 Nov 2013)-criteria for classification of hydraulic structures by hazard class.

Data availability

Data will be made available on request.

References

- [1] K.A. Hudson-Edwards, D. Kemp, L.A. Torres-Cruz, M.G. Macklin, P.A. Brewer, J.R. Owen, D.M. Franks, E. Marquis, C.J. Thomas, Tailings storage facilities, failures and disaster risk, *Nat. Rev. Earth Environ.* 5 (9) (2024) 612–630, <https://doi.org/10.1038/s43017-024-00576-4>.
- [2] *Global Tailings Review, Global Industry Standard on Tailings Management*, ICMM, 2020.
- [3] D. Kemp, J.R. Owen, É. Lèbre, Tailings facility failures in the global mining industry: will a 'transparency turn' drive change? *Bus. Strat. Environ.* 30 (1) (2021) 122–134, <https://doi.org/10.1002/bse.2613>.
- [4] L.N. Bowker, D.M. Chambers, *The Risk, Public Liability, and Economics of Tailings Storage Facility Failures*, 2015 [Report].
- [5] J.C. Santamarina, L.A. Torres-Cruz, R.C. Bachus, Why coal ash and tailings dam disasters occur, *Science* 364 (6440) (2019) 526–528, <https://doi.org/10.1126/science.aax1927>.
- [6] L. Picciullo, E.B. Storrøsten, Z. Liu, F. Nadim, S. Lacasse, A new look at the statistics of tailings dam failures, *Eng. Geol.* 303 (2022) 106657, <https://doi.org/10.1016/j.enggeo.2022.106657>.
- [7] D. Lumbruso, M. Davison, R. Body, G. Petkovšek, Modelling the Brumadinho tailings dam failure, the subsequent loss of life and how it could have been reduced, *Nat. Hazards Earth Syst. Sci.* 21 (1) (2021) 21–37, <https://doi.org/10.5194/nhess-21-21-2021>.
- [8] P.K. Robertson, L. de Melo, D.J. Williams, G.W. Wilson, *Report of the Expert Panel on the Technical Causes of the Failure of Feijão Dam I*, 2019 [Technical report].
- [9] L.H. Silva Rotta, E. Alcântara, E. Park, R.G. Negri, Y.N. Lin, N. Bernardo, T.S.G. Mendes, C.R. Souza Filho, The 2019 Brumadinho tailings dam collapse: possible cause and impacts of the worst human and environmental disaster in Brazil, *Int. J. Appl. Earth Obs. Geoinf.* 90 (2020) 102119, <https://doi.org/10.1016/j.jag.2020.102119>.

- [10] D.M. Franks, M. Stringer, L.A. Torres-Cruz, E. Baker, R. Valenta, K. Thygesen, A. Matthews, J. Howchin, S. Barrie, Tailings facility disclosures reveal stability risks, *Sci. Rep.* 11 (1) (2021) 5353, <https://doi.org/10.1038/s41598-021-84897-0>.
- [11] C. Roche, K. Thygesen, E. Baker, *Mine Tailings Storage: Safety is No Accident*. UN Environment, GRID-Arendal, 2017. <https://www.grida.no/publications/383>.
- [12] A. Hopkins, D. Kemp, *Credibility Crisis: Brumadinho and the Politics of Mining Industry Reform*, CCH Australia, 2021.
- [13] ANCOLD, *Guidelines on tailings dams: planning, design, construction, operation and closure*. Australian National Committee on Large Dams, second ed., 2012.
- [14] FEMA, *Federal Guidelines for Dam Safety: Hazard Potential Classification System for Dams*, U.S. Department of Homeland Security, 2004.
- [15] World Bank, *Good Practice Note on Dam Safety – Technical Note 7: Tailings Storage Facilities*, World Bank, Washington, DC, 2021.
- [16] W. Graham, *A Procedure for Estimating Loss of Life Caused by Dam Failure*, U.S. Department of Interior Bureau of Reclamation Dam Safety Office Denver, Colorado, 1999.
- [17] S.N. Jonkman, J.K. Vrijling, A.C.W.M. Vrouwenvelder, Methods for the estimation of loss of life due to floods: a literature review and a proposal for a new method, *Nat. Hazards* 46 (3) (2008) 353–389, <https://doi.org/10.1007/s11069-008-9227-5>.
- [18] USBR, *Reclamation Consequence Estimating Methodology*, U.S. Department of the Interior Bureau of Reclamation, 2015.
- [19] M. Rico, G. Benito, A.R. Salgueiro, A. Díez-Herrero, H.G. Pereira, Reported tailings dam failures, *J. Hazard Mater.* 152 (2) (2008) 846–852, <https://doi.org/10.1016/j.jhazmat.2007.07.050>.
- [20] P. Concha Larrauri, U. Lall, Tailings dams failures: updated statistical model for discharge volume and runout, *Environments* 5 (2) (2018) 28, <https://doi.org/10.3390/environments502028>.
- [21] M. Meadows, S. Jones, K. Reinke, Vertical accuracy assessment of freely available global DEMs (FABDEM, copernicus DEM, NASADEM, AW3D30 and SRTM) in flood-prone environments, *Int. J. Digit. Earth* 17 (1) (2024) 2308734, <https://doi.org/10.1080/17538947.2024.2308734>.
- [22] European Space Agency, *Copernicus DEM – Global and European Digital Elevation Model (GLO-30)*, 2021.
- [23] C.T. Lloyd, A. Sorichetta, A.J. Tatem, High resolution global gridded data for use in population studies, *Sci. Data* 4 (1) (2017) 170001, <https://doi.org/10.1038/sdata.2017.1>.
- [24] F.R. Stevens, A.E. Gaughan, C. Linard, A.J. Tatem, Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data, *PLoS One* 10 (2) (2015) e0107042, <https://doi.org/10.1371/journal.pone.0107042>.
- [25] S.N. Jonkman, J.K. Vrijling, Loss of life due to floods, *J. Flood Risk Manag.* 1 (1) (2008) 43–56, <https://doi.org/10.1111/j.1753-318X.2008.00006.x>.
- [26] T. Munroe, G. Tondapu, E. Anderson, R. Griffin, *Online dam inundation mapping tool: Model vs. reality*, Esri User Conference, San Diego, CA, USA (conference poster) (2017).
- [27] D. Yamazaki, K. Ikeshima, J. Sosa, P.D. Bates, G.H. Allen, T.M. Pavelsky, MERIT hydro: a high-resolution global hydrography map based on latest topography dataset, *Water Resour. Res.* 55 (6) (2019) 5053–5073, <https://doi.org/10.1029/2019WR024873>.
- [28] WorldPop, *WorldPop global project population data: estimated residential population per 100x100m grid square*. <https://www.worldpop.org>, 2020.
- [29] D. Zanaga, R. Van De Kerchove, W. De Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, L. Li, N.-E. Tsendbazar, O. Arino, *ESA WorldCover 10 m 2020 v100* (Version v100), Zenodo (2021), <https://doi.org/10.5281/ZENODO.5571936>.
- [30] J. Marešová, K. Gdulová, P. Pracná, D. Moravec, L. Gábor, J. Prošek, V. Barták, V. Moudrý, Applicability of data acquisition characteristics to the identification of local artefacts in global digital elevation models: Comparison of the copernicus and TanDEM-X DEMs, *Remote Sens.* 13 (19) (2021) 3931, <https://doi.org/10.3390/rs13193931>.
- [31] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. Borla Tridon, B. Bräutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber, B. Wessel, G. Krieger, M. Zink, A. Moreira, Generation and performance assessment of the global TanDEM-X digital elevation model, *ISPRS J. Photogrammetry Remote Sens.* 132 (2017) 119–139, <https://doi.org/10.1016/j.isprsjprs.2017.08.008>.
- [32] T. Dedring, V. Graw, K. Thygesen, A. Rienow, Validation of an empirical model with risk assessment functionalities to simulate and evaluate the tailings dam failure in Brumadinho, *Sustainability* 14 (11) (2022) 6681, <https://doi.org/10.3390/su14116681>.
- [33] N. Ghahramani, A. Mitchell, N.M. Rana, S. McDougall, S.G. Evans, W.A. Take, Tailings-flow runout analysis: examining the applicability of a semi-physical area–volume relationship using a novel database, *Nat. Hazards Earth Syst. Sci.* 20 (2020) 3425–3438, <https://doi.org/10.5194/nhess-20-3425-2020>.
- [34] G. Malgesini, P.J. Talling, A.J. Hogg, D. Armitage, A. Goater, F. Felletti, Quantitative analysis of submarine-flow deposit shape in the marnoso-arenacea formation: what is the signature of hindered settling from dense near-bed layers? *J. Sediment. Res.* 85 (2) (2015) 170–191, <https://doi.org/10.2110/jsr.2015.15>.
- [35] N. Ghahramani, H.J. Chen, D. Clohan, S. Liu, M. Llano-Serna, N.M. Rana, S. McDougall, S.G. Evans, W.A. Take, A benchmarking study of four numerical runout models for the simulation of tailings flows, *Sci. Total Environ.* 827 (2022) 154245, <https://doi.org/10.1016/j.scitotenv.2022.154245>.
- [36] W. Pollock, J. Wartman, Human vulnerability to landslides, *GeoHealth* 4 (10) (2020) e2020GH000287, <https://doi.org/10.1029/2020GH000287>.
- [37] J.C. Helton, F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliab. Eng. Syst. Saf.* 81 (1) (2003) 23–69, [https://doi.org/10.1016/S0951-8320\(03\)00058-9](https://doi.org/10.1016/S0951-8320(03)00058-9).
- [38] J.A. León, M. Ordaz, E. Haddad, I.F. Araújo, Risk caused by the propagation of earthquake losses through the economy, *Nat. Commun.* 13 (1) (2022) 2908, <https://doi.org/10.1038/s41467-022-30504-3>.
- [39] C. Acerbi, D. Tasche, On the coherence of expected shortfall, *J. Bank. Finance* 26 (7) (2002) 1487–1503, [https://doi.org/10.1016/S0378-4266\(02\)00283-2](https://doi.org/10.1016/S0378-4266(02)00283-2).
- [40] R.J. Hyndman, Y. Fan, Sample quantiles in statistical packages, *Am. Statistician* 50 (4) (1996) 361–365, <https://doi.org/10.1080/00031305.1996.10473566>.
- [41] S. Leyk, A.E. Gaughan, S.B. Adamo, A. De Sherbinin, D. Balk, S. Freire, A. Rose, F.R. Stevens, B. Blankespoor, C. Frye, J. Comenetz, A. Sorichetta, K. MacManus, L. Pistolesi, M. Levy, A.J. Tatem, M. Pesaresi, The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use, *Earth Syst. Sci. Data* 11 (3) (2019) 1385–1409, <https://doi.org/10.5194/essd-11-1385-2019>.