A risk assessment tool for tailings storage facilities

Journal:	<i>Canadian Geotechnical Journal</i>
Manuscript ID	cgj-2020-0329.R1
Manuscript Type:	Article
Date Submitted by the Author:	23-Nov-2020
Complete List of Authors:	Chovan, Karen; Enviro Integration Strategies, Julien, Michel; Agnico Eagle Mines Ltd, Environment & Critical Infrastructure Ingabire, Edouardine-Pascale; Agnico Eagle Mines Ltd, Environmental Management James, Michael; Agnico Eagle Mines Ltd, ; Agnico Eagle Mines Ltd, Masengo, Edouard; Agnico Eagle Mines Ltd, Environmental Management Lépine, Thomas; Agnico Eagle Mines Ltd, Environmental Management Lavoie, Pascal; Agnico Eagle Mines Ltd, Environmental Management
Keyword:	risk assessment, tailings management, probability of failure, level of engineering, factor of safety
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

A risk assessment tool for tailings storage facilities

(Manuscript No. cgj-2020-0329). A revised manuscript submitted to the Canadian Geotechnical Journal - date of resubmission: November 23, 2020

Karen M. Chovan¹, Michel R. Julien, Edouardine-Pascale Ingabire, Michael James, Édouard Masengo, Thomas Lépine, Pascal Lavoie

Karen M. Chovan, P.Eng., P.M.P.
Enviro Integration Strategies Inc., 822 Main Street, Saskatoon, Saskatchewan, Canada S7H 0K3
Tel. (306) 270-1430
kchovan@envirointegration.com

Michel R. Julien, Eng., Ph.D., FEIC
Agnico Eagle Mines Limited, 400 - 145 King Street East, Toronto, Ontario, Canada M5C 2Y7
Tel. (514) 244-5876
michel.julien@agnicoeagle.com

Edouardine-Pascale Ingabire, Jr Eng., M.A.Sc.
Agnico Eagle Mines Limited – Complexe minier LaRonde, 10200, route de Preissac, Rouyn-Noranda, Quebec, Canada J0Y 1C0
Tel. (819) 759-3700 x 4105284
edouardine.ingabire@agnicoeagle.com

Michael James, Eng., Ph.D.
Agnico Eagle Mines Limited, 400 - 145 King Street East, Toronto, Ontario, Canada M5C 2Y7
Tel. (819) 763-1089
michael.james@agnicoeagle.com

Édouard Masengo, Eng., Ph.D.
Agnico Eagle Mines Limited, 400 - 145 King Street East, Toronto, Ontario, Canada M5C 2Y7
Tel. (819) 856-0108
edouard.masengo@agnicoeagle.com

Thomas Lépine, Eng.
Agnico Eagle Mines Limited, 400 - 145 King Street East, Toronto, Ontario, Canada M5C 2Y7
Tel. (418) 473-8077
thomas.lepine@agnicoeagle.com

Pascal Lavoie
Agnico Eagle Mines Limited, 400 - 145 King Street East, Toronto, Ontario, Canada M5C 2Y7
Tel. (819) 277-0045
pascal.lavoie@agnicoeagle.com

¹ Corresponding author.

Abstract

The recent occurrence of several major failures of tailings storage facilities (TSF) has caused the mining industry to focus on significantly improving the engineering and management (design, construction, operation, and monitoring) of these structures to reduce their environmental impact. This effort is led by the Mining Association of Canada, which mandates the application of risk assessment in tailings management. Due to the very complex nature of TSF, such as phased design and construction, continuous operation, and evolving guidelines and practices over many years, the application of traditional risk assessment tools has limitations. A risk assessment tool specifically developed for TSF management is presented. This tool is based on the work of Silva et al. (2008) that relates the annual probability of failure to the factor of safety and the level of engineering. This relationship was modified to reflect current practice. The annual probability of failure was then combined with a consequence rating to produce a rational and quantifiable evaluation of risk. The risk assessment tool provides detailed information on the level of practice of a structure, the corresponding annual probability of failure as well as the associated risk. Validation of the tool included application to a recent well-documented failure.

Keywords: risk assessment, tailings management, probability of failure, level of engineering, factor of safety

Résumé

Les récentes ruptures de parc à résidus miniers ont mené l'industrie minière à améliorer l'ingénierie et la gestion (la conception, la construction, l'opération et l'entretien) de ces infrastructures pour réduire leurs impacts environnementaux. L'association minière du Canada demande à ses membres d'analyser les risques liés à la gestion des résidus. Les parcs à résidus sont conçus, construits et opérés sur plusieurs années et l'industrie minière doit s'adapter à l'évolution de la pratique. Ces défis font en sorte que l'application d'outils traditionnels d'analyse de risques a certaines limites. Cet article présente un outil d'analyse de risques spécifiquement développé pour les parcs à résidus se basant sur les travaux de Silva et al. (2008) qui font le lien entre la probabilité annuelle de rupture au facteur de sécurité et au niveau d'ingénierie. La méthodologie a été modifiée pour refléter l'état actuel de la pratique. La probabilité annuelle de rupture est combinée au niveau de conséquence afin de produire une évaluation rationnelle et quantifiable du risque. Cet outil procure de l'information détaillée sur le niveau d'ingénierie d'une infrastructure, la probabilité annuelle de rupture et sur le niveau de risque. L'outil a été validé en utilisant l'étude de cas d'une rupture récente bien documentée.

Mots clés : analyse de risque, gestion des résidus miniers, probabilité de rupture, niveau d'ingénierie, facteur de sécurité

Introduction

In 2004, the Mining Association of Canada (MAC) implemented the Towards Sustainable Mining[®] program (TSM). The program has been updated periodically, most recently following the tailings storage facility (TSF) failures at Mount Polley (2014) and Fundão (2015). The primary objective of TSM is to ensure the responsible management of the risks associated with mining, particularly those affecting the environment. Implementation of TSM at mining sites in Canada and elsewhere is mandatory for MAC members. The program is adaptable to a broad range of organizational structures, mining activities, and regulatory environments. MAC produced a Guide to the Management of Tailings Facilities (the Tailings Guide) (MAC 2019). The Tailings Guide is a core component of TSM and specifies:

- A solid and flexible governance policy;
- Utilisation of Best Available Technology and Best Available/Applicable Practice;
- The identification of critical controls and their implementation as performance monitoring indicators; and
- The application of a risk-based approach.

Larger mining companies operate numerous and very diverse TSF. This diversity stems from differences in orebodies, mineralogy, climate, environment, ore processing, engineering practice and regulations as well as the type of tailings (slurry, thickened or filtered).

This paper presents a risk assessment tool for the management of TSF. The tool includes an empirical method to estimate the annual probability of failure that is then combined with a quantitative evaluation of the consequences of failure to produce an objective and detailed assessment of the risk. It can be used to assess the risk of a TSF portfolio of a mining company,

including the identification of key elements of a specific facility that can be targeted to manage the risk. Incidentally, it also results in improved documentation and internal communication. The use of this tool allows the intent and requirements of the Guide to be integrated into the complex operational realities faced by mining companies. While this paper focuses on TSF, the approach has been adapted to other critical mining infrastructures that are similar to TSF in terms of risk (water management infrastructure, heap leach facilities, and rock storage facilities), with minor modifications.

Governance and risk assessment

Historically, the mining industry did not consider mine waste management as a core business activity, as are excavating rock or processing ore. However, in the last decade, there has been a significant shift. As mentioned by Julien (2016), mine waste management is now viewed as a core business activity and the mining industry is directly involved through increased in-house capabilities, involvement in the design and operation of waste management facilities, as well as improved linkage with production.

The Tailings Guide now mandates a governance policy that places responsibility for tailings management with the Owner, who must designate an Accountable Executive Officer. One of the primary responsibilities of the Accountable Executive Officer is the integration of a risk-based approach into tailings management.

The Tailings Guide also includes the notion of a risk-based approach as an important element of a well-functioning tailings management system. However, conventional risk analysis, where risk is expressed as a product of the probability of occurrence and the consequence of failure (International Electrotechnical Commission 2009), is particularly challenging for TSF where the

frequency of occurrence, or annual probability of failure, is often low to very low and consequences are generally high to extreme. Using this method, many TSF are rated as low to medium risk, making it difficult to differentiate where priorities might exist and to utilize the information in TSF management. This challenge has been described by others (Caldwell et al. 2015; Oboni et al. 2016).

Furthermore, conventional risk analysis does not support the level of sophistication needed to have a meaningful risk-based approach for complex and critical infrastructure such as TSF, in particular when design, construction and operation typically span many years. An additional challenge comes from the very strong influence of advanced geotechnical phenomena including static liquefaction, post-peak and post-liquefaction behavior, residual strength, etc., complicating the communication of the risks. Senior management and mine site personnel often do not have the geotechnical background to appreciate the subtleties of these phenomena and their interaction with the performance and stability of TSF.

Risk assessment serves two primary purposes: 1) the prioritization of maintenance, monitoring and improvements (e.g., stabilization) by risk level; and 2) a means of communicating to management the specific risks associated with individual TSF and the entire TSF portfolio. The next sections describe the development and application of a simple yet robust risk assessment tool, specifically designed for compiling information, then measuring, understanding and communicating detailed risk assessment of TSF.

The probability of failure

Using data from actual geotechnical engineering infrastructures that included dams, natural and constructed slopes, earth-retaining structures and tailings dams, Silva et al. (2008) developed empirical relationships between annual probability of failure (APF), factor of safety (FS) and Level

of Engineering (LOE). The FS used was that calculated for the static condition. The LOE was determined through systematic evaluation of the design (investigation, laboratory testing and analysis), construction as well as operation and monitoring. The LOE was then classified as I, II, III or IV, respectively indicating Best, Above Average, Average or Poor. Briefly, the LOE is based on a quantitative evaluation of the quality and thoroughness of the design and construction, as well as the operation and maintenance of the structure. A more complete description is provided below. Figure 1 presents the relationships established between FS and APF for the different LOE by Silva et al. (2008). The case histories considered are plotted on the figure and the LOE is represented by the four curves. For a given factor of safety, the APF increases with decreasing LOE. As an example, an embankment with an FS of 1.5 and an Average LOE (Level III) has an APF of 1×10^{-2} , while an embankment with an FS of 1.3 and a Best LOE (Level I) has an APF of 1×10^{-4} . Assuming 100-year life spans, the embankment with a FS of 1.5 has a 63% chance of failure and the embankment with a FS of 1.3 has a 1% change of failure. Such relationships are powerful tools for communicating with technical and non-technical professionals.

The framework developed by Silva et al. (2008) was selected as the basis for the risk assessment of tailings storage facilities for three primary reasons. It was developed from case histories including embankments that have similar construction methods and failure mechanisms as TSF. It includes the Level of Engineering which recognizes that the quality and degree of uncertainty in design, construction, operation and monitoring can directly influence the probability of failure for a given factor of safety. Though the state of practice, regulations and guidelines are strongly driven by a deterministic view of the factor of safety, using the annual probability of failure is a more palatable concept to non-experts. Regulations and guidelines implicitly consider the APF for precipitation

and seismic loadings, where the recurrence interval of an event (e.g., 1:10 000 years) is the inverse of its annual probability (e.g., 1×10^{-4}).

Other authors have also referenced the work of Silva et al. (2008) to consider uncertainty in the evaluation of TSF stability. Herza and Phillips (2017) cited Silva et al. (2008) in their work that recommends minimum factors of safety based on the consequences of failure and the uncertainties in engineering analysis. The recommended minimum FS values increase with increasing uncertainty and increasing level of consequence.

As noted, this methodology uses the static factor of safety. This is considered reasonable given that the knowledge base and experience for this condition are solid and well established. Although more complex failure mechanisms (e.g., strain softening) are not explicitly included, it is assumed that the Best LOE considers all of the potential failure mechanisms for a particular structure and that lower LOE may not have considered some of these mechanisms – increasing the degree of uncertainty and the APF.

In the context of risk assessment, failure is defined as an event that would trigger the assumed consequences. There must be agreement between the failure and the consequences. For example, an event that does not result in the release of tailings could be defined as a failure in a general context but might not initiate the consequences considered in the risk assessment. Such an event should not be considered in the risk assessment.

Modification of the Silva Method for tailings storage facilities

The method for calculating the APF developed by Silva et al. (2008), the Silva Method, was modified to account for the complexities of the phased design and construction, the continuous operation of TSF over many years, the observed performance, and the changes in engineering

practice. The method was also modified to facilitate computerized calculation. These modifications are described below.

Calculation of the annual probability of failure

In the development of the risk assessment tool for tailings storage facilities, minor adjustments were made to the Silva et al. (2008) APF-FS relationships to create straight-line interpretations in semi-logarithmic space while preserving its intent. Additionally, a minimum APF was selected for each LOE. For Levels I, II, III and IV, the APF can be calculated using Equation 1 and the values in Table 1. For the mid-level relationships (Ib, IIb and IIIb), the APF can be mathematically interpolated from adjacent levels.

$$(1) APF = MAX[a \cdot e^{b \cdot FS}, c]$$

where:

a and b are constants, presented in Table 1;

FS is the factor of safety under static conditions; and

c is the minimum APF as per Table 1.

These modified relationships, the mid-level interpretations, and the original Silva relationships are compared on Figure 2. As shown on Figure 2, the modifications had minor impact on the relationships, except for slightly increased conservatism in LOE I and II between FS values of 1.2 to 1.8. The mid-level relationships were created to reduce the gaps between the APF for different LOE, which exceeded two orders of magnitude. For example, for an infrastructure with an FS of 1.8, the APF decreases from 2×10^{-6} to 1×10^{-3} for an LOE II and III respectively.

To establish minimum APF values for the LOE, several risk management guides were examined, including the Federal Guidelines for Dam Safety Risk Management (Federal Emergency Management Association [FEMA], 2015), the Risk-Informed Decision-Making Guideline (Federal Energy Regulatory Commission [FERC] 2016) and the Safety of Dams Policy and Procedures (United States Army Core of Engineers [USACE] 2014). For LOE I, the minimum APF was set to 1×10^{-8} , the lowest probability of failure typically plotted on most risk tolerance charts for dams. For LOE IIb, the minimum APF is 1×10^{-6} , generally accepted for risks associated for commercial aviation (Lacasse and Nadim 1998), and the greatest APF generally considered tolerable for estimated loss of life over 1 000 persons. For LOE IIIb, the minimum APF is 1×10^{-4} , which is considered tolerable provided mitigation measures can be taken in accordance with “*As Low As Reasonably Practical*” principles. The minima for other levels were established using judgement. The use of minimum APF values assumes any facility can fail and that there is always a certain degree of uncertainty, which decreases with increasing Level of Engineering but is never null.

The Level of Engineering – The Level of Practice

In the Silva Method, the Level of Engineering is based on a quantitative evaluation of three categories: 1) Design, which has three subcategories - Investigation, Laboratory Testing and Analysis and Documentation; 2) Construction; and 3) Operation and Monitoring, collectively referred to as DCO. Each of these categories and subcategories is evaluated using a number of criteria that provide points. These points are summed and factored (weighted) and the score is used to determine the LOE where scores of 1.0, 2.0, 3.0 and 4.0 correspond to Levels of Engineering of I (Best), II (Above Average), III (Average) and IV (Poor), respectively. The weighting is such that Design accounts for 60% of the score, with Investigation, Laboratory Testing, and Analysis and Document each being 20%. Construction and Operation and Monitoring both account for 20%.

In reviewing the Silva Method, it was felt that it could be improved and extended to consider management and review systems that are now integrated in engineering practice, while respecting the original intent. This generalization rests on four considerations for tailings storage facilities that may contribute towards a higher level of engineering:

1. Strong management systems;
2. Collaborative DCO, e.g., with good and continuous interaction between necessary stakeholders as proposed by MAC (2019): the Design Engineer, Responsible Person, Engineer of Record (EOR) and Independent Review Board;
3. Formal and thorough review processes, due to the higher likelihood of identification of technical challenges; and
4. Document management systems in place to ensure the retention of DCO information with time.

The criteria were modified by Julien et al. (2019) to include these considerations. In view of these modifications, the Level of Engineering is now referred to as the Level of Practice (LOP).

During subsequent development of the risk assessment tool, criteria associated with observed performance were added to the Operation and Monitoring category. These criteria consider the behavior of the TSF based on observations and monitoring, with the objective of including aspects such as unanticipated settlement, deformation or excessive seepage in the evaluation of the LOP.

The modified DCO criteria developed for the TSF risk assessment tool are presented on Table 2. The table includes descriptions for the ratings of each criterion. The criteria are rated based on the current state of the practice, which may differ from that at the time of design and construction. An LOP of I (High) requires that the infrastructure meets the requirements of the current state of the

practice from inception until now. Deviations or omissions from the current state of the practice result in a lower Level of Practice and thus an increased Annual Probability of Failure.

The Silva Method is mostly based on conventional geotechnical structures where there are distinct Design, Construction, and Operation and Monitoring phases. As TSF are generally designed, constructed, and operated over many years, the weighting of each category was adjusted accordingly. Julien et al. (2019) initially proposed that the weighting of Laboratory Testing be lowered from 20% to 10% and increasing Operation and Monitoring to 30%. This adjustment aligns with the trend in practice to rely more on field monitoring such as with piezometers, inclinometers, etc.

Since then, rather than assigning a fixed weighting to each category, it was decided to assign equal weighting to each criterion considered in the evaluation of the Level of Practice. Table 3 compares the corresponding weightings of each category and number of criteria from the Silva Method and Julien et al. (2019) with those presented here. From the Silva Method to the work presented here, there was an increase in the number of criteria for every category, with larger increases for Construction (from 3 to 8) and for Operation and Monitoring (from 3 to 14). With the criteria now being given equal weight, the weighting of the categories is determined by the number of criteria in each category. The weight of the Investigation category remained the same at 20%. The differences from the weighting of the Silva Method are that Laboratory Testing has changed from 20% to 13%, both Analysis and Documentation and Construction are now 18% rather than 20%, and Operation and Monitoring has increased from 20% to 31%. These new weightings were deemed suitable for the risk assessment of tailings storage facilities due to their ability to capture and adequately assess the impacts of the categories on the annual probability of failure.

Application of the modified method for tailings storage facilities

The modified method for the calculation of the Annual Probability of Failure of a tailings storage facility is described in the following paragraphs.

The consequences associated with the failure of the TSF are usually determined at critical cross-sections. Where the consequences of the failure of different cross-sections of a TSF are different, this should be noted and the APF calculations conducted for each cross-section. Information from tailings dam breach analysis, slope failure run-out analyses, or dam classification can be used to complete this task.

The Factor of Safety for the cross-sections being utilized in the analysis are those associated with static conditions in drained or undrained conditions, whichever is relevant for the section. If the TSF retains a pond, ranges in operating water levels are considered.

The criteria are evaluated individually and rated as level I, II, III, IV as per Table 2. For consistent evaluation of different types of TSF, such as for filtered-stack tailings or in-pit containment, if a criterion is found not to be applicable, then it is rated as “Not Applicable”. The ratings are converted to coefficients, where I is $1/NC$, II is $2/NC$, III is $3/NC$, IV is $4/NC$, and NC is the total number of criteria. The average of the scores for each DCO category are calculated. When a criterion is rated “Not Applicable”, it is scored using the average of the remaining criteria within the respective DCO category.

The individual scores for all of the criteria are summed and a Level of Practice is assigned as per Table 2. The lowest and highest possible scores are 1.0, indicating a High LOP (Level I) and 4.0, indicating a Low LOP (Level IV).

Using the score of the LOP and the FS, the Annual Probability of Failure is calculated from Equation 1 above.

Risk assessment of a sample TSF portfolio

The risk assessment tool was revised and refined by conducting a preliminary risk assessment of part of the TSF portfolio of a large mining company. The data and results presented herein are preliminary as some of the results are in the process of being verified with site personnel.

The sample TSF portfolio

The portfolio includes a diverse assortment of tailings storage facilities in terms of mine type (e.g., open pit or underground), ore mineralogy, ore processing, tailings deposition, tailings types (slurry, thickened, filtered), climate (arid to subarctic), regulatory jurisdiction, and configuration and age (a few years to decades). Thus, they have been designed, constructed, and operated in different and evolving states of practice and regulations.

Methodology

The methodology used in the risk assessment for each infrastructure section was straightforward:

1. Derivation of the LOP and the APF using the modified method for TSF, as presented above;
2. Determination of the Consequence Rating; and
3. Estimation of the Risk Rating.

Details for the determination of the Consequence Rating and the Risk Rating are provided below. For convenience, the risk assessment tool was developed in spreadsheet format.

Determination of the Level of Practice and the Annual Probability of Failure

For each infrastructure considered, the criteria presented in Table 2 were rated by the Engineer of Record for that site and an external auditor. The EOR had access to the available documentation for all lifecycle phases yet was not directly involved in the preparation of these documents, removing a potential source of bias. The ratings were not based on the knowledge of the EOR, but rather on documented evidence that was referenced in notes by the EOR presenting the rationale for each rating.

DCO category scores were calculated from the criteria ratings and the corresponding LOP was then determined using Table 4. The static Factor of Safety was obtained from the most recent stability analysis of the TSF by the Design Engineer that had been reviewed by the EOR. Finally, the Annual Probability of Failure was calculated using Equation 1.

Lastly, the results were reviewed with representatives of the site, including the Responsible Person, to verify the ratings that were given, and to provide an opportunity for the representatives to reveal any additional documentation required for the evaluation. Notably, some of the ratings were improved over the original evaluations, affecting the overall DCO score and the resulting Annual Probability of Failure.

Determination of Risk Rating

The consequences of failure for each TSF, or more specifically each cross-section, were determined by tailings dam breach analysis, slope failure run-out analysis or other appropriate methods. Generally, this was conducted by the Design Engineer with review by the Engineer of Record and the Independent Review Board. The details are not presented here.

The consequences were then used to ascertain the Consequence Rating for each TSF. These ratings were based on a classification chart typical of that used in the mining industry. It is similar to the

dam classification system of the Canadian Dam Association (CDA) (2013) in that it considers human health and safety, environmental loss, damage to public infrastructure, and economic losses to other parties. However, unlike CDA (2013), it also considers effects on local communities and damage to the mining company (reputation and financial). Generally, these rating systems are slightly more conservative than CDA (2013). The Consequence Ratings are: Negligible (a Consequence Rating value of 1), Low (2), Medium (3), Major (4) and Extreme/Critical (5).

For the risk assessment of the Sample TSF portfolio, the Probability Ratings were assigned a value from 0, indicating Negligible, to 5, indicating Very High, based on the APF calculated from the DCO score and the FS. The conversion from APF to the Probability Rating is presented in Table 5. The ranges of APF for the Probability Ratings in the table are a result of the rationale presented above.

Risk is a product of probability and consequence. The Risk Rating, RR, is thus:

$$(2) \ RR = CR \cdot PR$$

where CR is the Consequence Rating and PR is the Probability Rating.

Risk ratings and risk classifications were obtained from the chart in Table 6 (adapted from Agnico Eagle Mines, 2020 and provided as an example). Risks were evaluated as Very High (for Risk Rating values of 20 to 25), High (12 to 19), Medium (5 to 11), Low (1 to 4), and Negligible (0). Figure 3 is a visual representation of Table 6, in which the limits of each risk classification are identified on a graph which relates the APF to the Consequence Rating.

Summary of results

As noted, this is a preliminary evaluation that was used to validate and refine the risk assessment tool. For convenience, the risk assessment was completed for the most critical cross-section of each infrastructure in the sample TSF portfolio.

Levels of Practice

The detailed results of the LOP scoring for the sample TSF portfolio are summarized on Table 7 and shown graphically on Figure 4. In Table 7 and Figure 4, “TEX” refers to an example tailings embankment cross section in the sample portfolio.

As noted, the scores can range from 1.0 to 4.0, with 1.0 being Level I (High LOP) and 4.0 being Level IV (Low LOP). For this portfolio the LOP scores vary from 1.1 to 2.6, indicating Levels from I (High LOP) to IIb (between Good/Average LOP). None of the infrastructure was rated Average or Poor. The variation in the LOP shown on Figure 4 is the result of differences in regional practice as well as changes in practice with time.

Recall that the LOP score is the sum of scores for the five DCO categories. This is illustrated in Figure 4, which presents the breakdowns of the LOP scores. Here, each category is identified by a different segment and the lengths of the segments are indicative of the category scoring. As a higher score is indicative of a lower LOP level, a longer bar length indicates poorer performance and/or higher uncertainty. This figure allows comparison of the overall LOP as well as the contribution of each category. For example, TEX 29 shows an LOP between Good and Average (Level IIb) while TEX 25 to TEX 28 are rated at Good LOP (Level II) or better. The detailed results show that, for TEX 29, the rating is lower for Investigation and Laboratory Testing, i.e., the bar lengths are longer for these categories compared to the other highlighted infrastructures. During the evaluation of the LOP for TEX 29, it was found that documentation management was insufficient, which revealed

uncertainties about the actual condition of the dam, leading to an increased Annual Probability of Failure.

Annual Probabilities of Failure

The APF for the sample TSF portfolio are listed in Table 7 and shown graphically on Figure 5 which presents the APF as a function of the Factor of Safety for each TSF considered. The APF varies from about 1×10^{-3} to 1×10^{-8} , a fairly extensive range and significant distribution. The variation of the APF is the result of variation of the Factory of Safety (from 1.4 to 2.8), but just as significantly, the variation of the LOP (from Level I to Level IIb).

Risk Ratings and Risk Classification

The Consequence Ratings, Probability Ratings, and resulting Risk Ratings for the TSF portfolio are given in Table 7. The Risk Ratings are shown on Figure 6, which permits a classification of risk across the portfolio. Figure 6 shows that 8 structures are classified as High Risk, 10 as Medium Risk, 5 as Low Risk, and 16 as Negligible Risk.

Discussion of results

The results of the preliminary risk assessment conducted on the sample TSF portfolio are discussed below. To broaden the applicability as well as the parameters of the discussion it is framed in general terms with some examples drawn from the results of the sample TSF portfolio.

Risk tolerance level

Application of risk assessment is facilitated by the identification of the tolerable level of risk of an organization. This can be expressed as a probability of failure, on an annual basis or on an expected lifespan, as a recurrence interval or as a Risk Rating. Note that this exercise is independent of the selection of design criteria. When a probability of failure or recurrence interval is used, the

consequences of failure are not considered, as they are with the Risk Rating. For purposes of discussion, an APF of 1×10^{-4} , corresponding to a recurrence interval of 1:10 000 years, was selected as a tolerable occurrence rate for a failure event. This level aligns with the societal tolerable risk limit provided risk mitigation measures are taken in accordance with “*As Low As Reasonably Practical*” principles (FEMA, 2015; FERC, 2016; USACE, 2014).

General

The primary purpose of a risk assessment is to provide quantitative, comprehensible and actionable measures of the risks associated with the infrastructure evaluated. This risk assessment tool does that and it provides useful, understandable and communicable data and visuals for a range of metrics associated with the Design, Construction Management and QA/QC, and Operation and Monitoring of TSF. The results also provide insights into where improvement efforts might be focused, as the scores reflect both actual conditions, where accurate information is available, and uncertainties, where information or documentation is lacking.

It has been noted that the Factor of Safety, though commonly used as a deterministic metric of stability, may include a considerable degree of uncertainty. Merely achieving a minimum FS does not assure stability. By combining the FS with a rational evaluation of the LOP, an APF can be calculated which incorporates the degree of uncertainty and the rigor of engineering, project management, and review. Additionally, the APF can be combined with an analysis of the consequences of failure to produce a value of risk. An organization can use this information to manage a TSF portfolio in one of three ways:

1. Establish an allowable occurrence rate for a failure event for all infrastructure, regardless of consequences (e.g., an APF of 1×10^{-4} , which is equivalent to a recurrence interval of 1:10 000 years);
2. Set a maximum allowable Risk Rating (e.g., Medium Risk); or
3. Combine a maximum allowable APF (e.g., 1×10^{-3}) with a maximum Risk Rating (e.g., Medium Risk).

The first option, because it does not consider consequences, is very conservative for infrastructures of low consequences. For example, with a maximum allowable APF of 1×10^{-4} , the minimum allowable FS for TSF rated at LOP of I, II and III are, respectively, 1.3, 1.5 and 2.1. From Figure 2, an LOP of IV is not possible for a maximum allowable APF of 1×10^{-4} . With the second option, there is recognition that significant consequences of failure should be associated with a lower APF and a higher LOP. However, it may also allow a high APF when the consequences are low. For example, referring to Figure 6, a debate could arise from tolerating an infrastructure with Consequence Ratings of 1 or 2 (Negligible to Minor) that has an APF of 0.01 (a recurrence interval of 1:100 years). The third option considers the consequences of failure and by setting a maximum allowable APF does not accept that infrastructure of low consequence could have a relatively high rate of failure.

High DCO scores can indicate an overall poor level of practice with time, poor practice in a particular DCO category, or categories, or simply insufficient documentation (e.g., missing construction as-built drawings and high uncertainty with respect to construction). The LOP of an infrastructure is not a constant. It evolves during its lifespan and it can be improved. For example, during the preliminary risk assessment, interaction between the EOR and site personnel resulted in the discovery of construction as-built documents that resulted in a direct improvement of the LOP

scores for some infrastructures. In another case, fieldwork that was completed just before the assessment resulted in an improved LOP score when the geotechnical report was received, and some uncertainties were reduced.

The tool also allows the determination of a targeted minimum FS to achieve a desired APF. This targeted minimum FS may be helpful in selecting and designing mitigation measures. Assume that the maximum allowable APF was set at 1×10^{-4} . An infrastructure with an LOP of IV (Poor) would never meet the objective, no matter the Factor of Safety. Work to improve the LOP would be required. Figure 5 shows a few structures within the sample portfolio that did not satisfy the assumed maximum allowable APF. The DCO rating for these structures varies from 1.6 to 2.6, and thus, the LOP varies between II (Good) and IIb (between Good and Average). The DCO scores of these structures could be improved by works such as additional site investigation and laboratory testing, increased monitoring, and updated stability analyses, which would result in a lower APF. If these efforts were not sufficient in attaining the maximum allowable APF, then some means of stabilization (e.g., buttressing, foundation improvement, drainage, lowering of the pond level) could be considered to increase the FS and thus lower the APF.

The detailed breakdown of DCO scores on Figure 4 could be used to target where improvements might be made. Taking TEX 29 as an example, the bar length for Investigation, shown in black, is significantly longer than that of TEX 25. As discussed above, the documentation of the investigations was lacking. One way to improve the score for TEX 29 would be to complete additional site investigations to remove these uncertainties. Similarly, the bars in different shades of grey for TEX 29 indicate additional Laboratory Testing and Analysis and Documentation would also be worth considering.

Validation of the tool with the Mount Polley TSF failure

Validation of the risk assessment tool included applying it to well-documented case studies of TSF failures. Since the Silva Method is empirically based on such studies, this an appropriate method of validation. As an example, the risk assessment tool was applied to the 2014 failure of the Mount Polley Mine TSF. This validation is in no way to provide an opinion on this failure but simply to test the tool. The information used in the risk assessment was taken from reports issued in 2015 by the Chief Inspector of Mines (British Columbia Ministry of Energy and Mines [BCMEm] 2015) and the Independent Expert Engineering Investigation and Review Panel (British Columbia 2015).

The Mount Polley Mine is located in south-central British Columbia, Canada, where slurry tailings were produced and deposited, along with runoff water, in an incrementally raised tailings impoundment. We understand that the impoundment was retained by three dams, the Main Embankment, the Perimeter Embankment and the South Embankment, all of which were constructed using the modified centerline method. At the time of failure, the TSF had reached a height up to 50 m. Figure 7 is a partial plan of the TSF with the locations of the cross-sections considered in the evaluation.

Prior to the failure, the Main Embankment had been considered the most critical structure of the TSF in terms of stability, due to the presence of strain-softening glaciolacustrine material in its foundation and displacements observed in this material during construction. However, the failure occurred at the Perimeter Embankment and was apparently linked to the presence of a glaciolacustrine unit in its foundation not identified during all previous geotechnical investigations.

Annual Probability of Failure evaluation for assumed conditions pre-failure

Pre-failure, the static FS of the Perimeter Embankment and Main Embankment were calculated to be 1.5 and 1.4, respectively (BCMEM, 2015). The LOP of the Perimeter Embankment and Main Embankment were evaluated by the authors using the references provided above and were determined to be III (Average) and IIb (Average/Good), respectively.

The better LOP for the Main Embankment can be explained by the additional laboratory testing, analyses, and monitoring that was done due to the challenges related to the observed displacements of the foundation unit during construction. The corresponding values of APF for the Perimeter Embankment and Main Embankment were estimated to be, respectively, 1×10^{-2} and 4×10^{-3} . Thus, although the Perimeter Embankment had a higher FS, it had a lower LOP and a higher APF than the Main Embankment. This is shown on Figure 8a. Figure 8b shows the risk category ratings for both embankments using the Consequence Rating of 5 (Extreme).

Comparison to post-failure investigations

Extensive, post-failure investigations revealed the presence of a glaciolacustrine unit in the foundation of the Perimeter Embankment that was not noted in previous investigations. We understand that the presence of this material ultimately led to the August 2014 failure as it had lower strength than that considered in the stability analyses for design. Additional stability analyses

of the Perimeter Embankment were conducted to include the findings of the post-failure investigations. As part of the post-failure investigations, the critical cross-section for the Perimeter Embankment was cross-section C where the failure occurred. The additional data points presented in Figure 8 assume no improvement in the LOP due to the post-failure investigations, and correspond to the following cases (BCMEM 2015):

- A FS of 1.3 was obtained considering peak-drained strength in the glaciolacustrine unit and steady state pore pressures;
- A FS of 1.2 was obtained considering peak-drained strength in the glaciolacustrine unit and excess pore water pressure generated by the raising of the embankment; and
- A FS of 1.0 was obtained considering undrained strength in the glaciolacustrine unit.

The post-failure investigations and analyses indicated that there were significant uncertainties in the stratigraphy and the foundation strength parameters of the Perimeter Embankment, each having an impact on the reliability of the calculated FS. Indeed, it is interesting to note that the Perimeter Embankment fell within the Very High-Risk category, although the calculated pre-failure FS respected applicable guidelines for static conditions (CDA 2013). This highlights the risk of relying solely on the FS as a metric for stability, and, more importantly, the value of a rigorous and quantitative risk assessment. Finally, the intent of this exercise was to test the tool against a well-documented case. It showed the tool could be used to assess the risk profile and associated uncertainties of a complex system beyond a purely deterministic assessment based on FS.

Discussion

There are a few other topics that should be discussed with respect to the risk assessment tool, such as the effect of different types of tailings, the relationship between risk assessment and reliability

index methods, factoring the Factor of Safety based on the Level of Practice, implications of the life span of a structure, the necessary qualifications of risk assessment personnel and avoiding having risk assessment become a routine check-box exercise.

Although the use of slurry or thickened tailings remains widespread, the use of filtered tailings is increasing. The risk assessment tool described here can, and has been, applied to existing TSF containing filtered tailings. While these examples were not specifically identified to highlight the differences, we can confirm that the LOP evaluation does not necessarily favor filtered over slurry cases. In fact, the LOP and APF of the filtered tailings TSF do fall within the average range of the full set. The risk rating is most affected by the presumed lower consequences of a filter tailings stack failure if these materials are shown to be non-liquefiable.

Reliability-based design is common in general civil engineering practice (Fenton et al. 2016), but much less so in geotechnical engineering, particularly for the mining industry. The basis of reliability-based design is the consideration of uncertainties (in material properties, material models, methods of analysis and loadings) and of heterogeneity (in material properties) by factoring the parameters used in the calculation of the FS. These factors are estimated using statistical analysis or from published sources. The resulting FS maybe a singular value corresponding to a given probability (or confidence interval) or a probabilistic distribution. The risk assessment tool presented here is different in that the uncertainties considered are those associated with the Level of Practice applied to determine parameters used in the calculation of the FS as well as many other aspects of design, construction, operation and monitoring that affect the probability of failure. That being said, the two methods are not mutually exclusive and could both be applied to an infrastructure.

Despite its inherent limitations, the Factor of Safety will continue to be used in engineering practice. Another possible use for the risk assessment tool is the establishment of minimum FS values for specific Levels of Practice to achieve a targeted APF. For example, minimum FS values of 1.28, 1.5, 2.07 and 2.82 for LOP of I, II, III and IV, respectively, each result in an APF of 1.32×10^{-4} (per Equation 3, a rearrangement of Equation 1). Alternatively, given the calculated FS, LOP and APF for a structure, Equation 1 (or Figure 2) could be used to calculate an equivalent FS corresponding to a specific LOP or APF. For example, if a structure has a FS of 2.5, an LOP of III, and an APF of 5.31×10^{-6} (per Equation 1), the equivalent FS for a LOP of II is 1.73 (with the same APF). So, a mining company or a regulatory agency could establish a standard LOP, II for example, and establish a minimum equivalent FS.

$$(3) FS = \frac{\ln(APF/a)}{b}$$

Though most TSF are permanent infrastructure with indefinite lifespans, there is a hopeful trend towards landform development at closure that would eliminate the possibility of failure. Some TSF and most water management infrastructure have lifespans that can be reliably estimated. In these cases, the lifespan probability of failure may be a useful parameter to calculate from the APF. This can be done with Equation 4. In these cases, the Probability Ratings (Table 5) should be revised based on owner tolerance.

$$(4) LPF = 1 - (1 - APF)^{N_L}$$

where LPF is the probability of failure during a lifespan of N_L years.

Evaluation of the criteria for determination of the Level of Practice requires extensive knowledge of the geotechnics of mine waste, mine waste management and water management as well as good

knowledge of the site, its operation and history. In the risk assessment for the example TSF portfolio presented here, the criteria and LOP were evaluated by experienced Engineers of Record, each having qualifications aligned with CDA (2019), with the collaboration of a competent external auditor.

As with any standardized evaluation system, there is potential that the risk assessment tool could become a “check-box” type exercise in some organizations, diminishing the usefulness of the results. To avoid this, safeguards could be implemented, such as: 1) be done by a qualified professional (e.g. Engineer of Record) partnered with an external auditor; 2) include notes in the evaluation record for each criterion score with rationale and references to documents (e.g. technical reports and plans); 3) be formally reviewed with site personnel, including the Responsible Person; 4) be subject to review by the Independent Review Board and the Accountable Executive Officer; and 5) be integrated and linked to the Operation, Maintenance and Surveillance Manual or the Tailings Management System.

Conclusion

A risk assessment tool for TSF was developed from the work of Silva et al. (2008). The tool explicitly integrates tailings management systems as recommended by the Mining Association of Canada in its Tailings Guide. The tool was used to conduct a risk assessment of a diverse sample TSF portfolio. It was also validated by applying it to the case study of the Mount Polley TSF failure, where the evaluation showed that the tool is powerful in clearly highlighting risks associated with complex critical infrastructure, such as TSF. Thus, it is well suited for tailings management. The tool is being adapted for other critical infrastructure, including water management infrastructure, heap leach facilities, and rockfill storage facilities.

While risk assessment typically includes some degree of subjectivity, the proposed approach is quite objective in that it is based on answers to quantitative queries that must be supported by documentation. Furthermore, by including review and collaboration, objectivity is further strengthened. Conducting the evaluation of the Level of Practice first with the Engineer of Record and then reviewing the results with site personnel, including the Responsible Person, facilitated the process – where information was lacking and scores were low, site personnel were quick to provide any additional documentation to improve the scores. The site operators also appreciated having transparency of a list of criteria upon which TSF performance is evaluated, enhancing their ability to support TSF management.

The risk assessment tool provides a detailed understanding of the risks associated with a TSF portfolio as well as indicators that can be used to prioritize investigation, monitoring and corrective actions. Furthermore, by considering the Level of Practice, the engineering, operation, monitoring, and performance of TSF, as well as the factor of safety, are directly implicated in discussions of stability. The use of such risk assessment tools should facilitate risks communication with the management, stakeholders and any other non-geotechnical people, and identification of priorities for risk mitigation.

Acknowledgements

The authors appreciate the support and information provided by Agnico Eagle Mines Ltd. that made the development of this risk assessment tool possible as well as Professor Francisco Silva and his colleagues who conducted the research that formed the basis of the methodology used in the tool.

References

- Agnico Eagle Mines Ltd. 2020. HSEC Risk Analysis Quick Reference Card.
- British Columbia. 2015. Independent Expert Engineering Investigation and Review Panel. Report on Mount Polley tailings storage facility breach. British Columbia, Canada: Province of British Columbia.
- British Columbia Ministry of Energy and Mines. 2015. Investigation report of the Chief Inspector of Mines - Mount Polley Mine tailings storage facility breach, August 4, 2014. British Columbia, Canada: Ministry of Energy and Mines - Mining and Mineral Resources Division.
- Caldwell, A. J., Oboni, F., and Oboni, C. 2015. Tailings facility failures in 2014 and an update on failure statistics. *In* Tailings and Mine Waste 2015, October 26-28, 2015, Vancouver, Canada.
- Canadian Dam Association. 2013. Dam safety guidelines 2007 (2013 ed.). Toronto, Canada: Canadian Dam Association.
- Canadian Dam Association. 2019. Technical Bulletin: Application of Dam Safety Guidelines to Mining Dams (2019 ed.). Toronto, Canada: Canadian Dam Association.
- Federal Emergency Management Association. 2015. Federal guidelines for dam safety risk management (No. FEMA P-1025). USA: Federal Emergency Management Association.
- Federal Energy Regulatory Commission. 2016. Chapter 3 - Risk assessment. Risk-informed decision making (RIDM) - Risk guidelines for dam safety (Version 4.1). pp. 3-1-3-42. USA: Federal Energy Regulatory Commission.

- Fenton, G.A., Naghibi, F., Dundas, D., Bathurst, R.J., and Griffiths, D.V. 2016. Reliability-based geotechnical design in 2014 Canadian Highway Bridge Design Code. *Canadian Geotechnical Journal*, 53, 236-251.
- Herza, J., and Phillips, J. 2017. Design of dams for mining industry. In Annual Meeting of the International Conference on Large Dams Conference, July 3-7, 2017, Prague, Czech Republic.
- International Electrotechnical Commission. 2009. Risk management - Risk assessment techniques (No. IEC 31010:2009). Geneva, Switzerland: IEC.
- Julien, R. M. 2016. Conference presentation. "Governance aspects with mine waste management facilities". In 24th World Mining Congress: Mining in a World of Innovation, October 18-21, 2016, Rio De Janeiro, Brazil.
- Julien, R. M., Masengo, E., Lavoie, P., and Lépine, T. 2019. A risk-based approach to support the implementation of the new guidelines related to tailings management. In Tailings 2019: 6th International Seminar on Tailings Management, July 10-12, 2019, Santiago, Chile.
- Lacasse, S., and Nadim, F. 1998. Risk and reliability in geotechnical engineering. In 4th International Conference on Case Histories in Geotechnical Engineering, March 8-15, 1998, St-Louis, USA, pp. 1172-1192.
- Mining Association of Canada. 2019. A guide to the management of tailings facilities (Version 3.1). Ottawa, Canada: Mining Association of Canada.

Oboni, F., and Oboni, C. 2016. A systemic look at tailings dams failure process. *In* Tailings and Mine Waste 2016, October 2-5, 2016, Keystone, USA.

Silva, F., Lambe, T. W., and Marr, W. A. 2008. Probability and risk of slope failure. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(12), 1691-1699. doi: 10.1061/(ASCE)1090-0241(2008)134:12(1691).

United States Army Corps of Engineers. 2014. Safety of dams - Policy and procedures (No. ER 1110-2-1156). Washington, DC, USA: Department of the Army, US Army Corps of Engineers.

Draft

List of acronyms

APF	Annual Probability of Failure
BCMEM	British Columbia Ministry of Energy and Mines
CR	Consequence Rating
DCO	Design, Construction, Operations and Monitoring
EOR	Engineer of Record
FEMA	Federal Emergency Management Association
FERC	Federal Energy Regulatory Commission
FS	Factor of Safety
LOE	Level of Engineering
LOP	Level of Practice
LPF	Lifespan Probability of Failure
MAC	Mining Association of Canada
PR	Probability Rating
RR	Risk Rating
TSF	Tailings Storage Facility
TSM	Towards Sustainable Mining
USACE	United States Army Corps of Engineering

List of tables

Table 1. Constants and minimum values for calculation of the Annual Probability of Failure.

Table 2. List of Level of Practice criteria.

Table 3. Number of Level of Practice criteria and weighting for Silva et al. (2008), Julien et al. (2019) and the current work.

Table 4. Level of Practice for DCO Score.

Table 5. Probability ratings.

Table 6. Risk classification chart (adapted from Agnico Eagle Mines Ltd., 2020).

Table 7. Detailed results of risk assessment of the sample TSF portfolio.

Draft

List of figures

Figure 1. Empirical relationships for annual probability of failure vs factor of safety for different levels of engineering (adapted from Silva et al. 2008).

Figure 2. Comparison of annual probability of failure – factor of safety relationships from the Silva Method and the proposed method.

Figure 3: Example risk classifications chart.

Figure 4. DCO scores and level of practice of sample TSF portfolio.

Figure 5. Estimated annual probabilities of failure of sample TSF portfolio.

Figure 6. Risk classification chart with results of sample TSF portfolio.

Figure 7. Plan view of the Mount Polley tailings storage facility showing location of cross-sections evaluated for the case study (adapted from BCMEM 2015).

Figure 8. Results of risk assessment of the Mount Polley tailings storage facility: a) annual probability of failure; and b) risk classification.

Table 1. Constants and minimum values for calculation of the Annual Probability of Failure.

Level of Practice, LOP	Constant, a	Constant, b	Minimum APF, c
I	21,100	-14.81	1×10^{-8}
II	143,040	-13.87	1×10^{-7}
III	661.66	-7.456	1×10^{-5}
IV	129.67	-4.872	1×10^{-4}

Draft

Table 2. List of Level of Practice criteria.

Criteria	Level Descriptions
Design - Investigation	
1.1 Baseline and background information collected about infrastructure-specific siting	I - Very detailed II - Detailed, may include some uncertainties III - Considered incomplete to fully understand the site IV - Limited, incomplete or containing significant gaps
1.2 Investigation program development, and adjustments to fill gaps, in relation to design, construction and operation phases	I - Highly connected, adapted and updated as necessary to address gaps, with thorough collaboration and review II - Some linkage, some gaps still exist, with regular collaboration and review III - Limited linkage, limited collaboration and review IV - Limited linkage, non-collaborative and without review
1.3 Investigation QA/QC program	I - Comprehensive program to verify quality/validity of results with limited uncertainties or gaps II - Good program, some uncertainties and gaps still exist III - Average program, significant uncertainties and gaps exist IV - Weak program
1.4 Current understanding of soil profile, stratigraphy, and geology for infrastructure-specific foundations and embankment conditions	I - High level of certainty II - Good level of certainty, some gaps may exist III - Average to good level of certainty, gaps regarding lateral heterogeneity IV - Low level of certainty
1.5 Current understanding of foundational strength characteristics, particularly for weak and soft zones, and level of reliance on in situ information	I - High level of certainty, heavy reliance on in situ information II - Good level of certainty, reliance on in situ information, some gaps III - Average to good level of certainty, limited in situ information IV - Low level of certainty, particularly for weak zones
1.6 Current understanding of tailings or other contained materials' mechanical behaviour, based on in situ characteristics and strength	I - Very detailed in situ information II - Reliable in situ information III - Limited in situ information IV - Assumed from literature and experience
1.7 Current understanding of groundwater conditions and piezometry, and basis on field data through time	I - High level of certainty through time, robust, reliable data II - Good level of certainty through time, robust program III - Approximate information, limited measurements over time IV - Low level of certainty through time

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
1.8 Robustness of hydrology assessments, including catchment area and site water balance, and basis on relevant field data	<p>I - Highly robust models based on field measurements and calibration with local stations for more than 10 years</p> <p>II - Models based on adequate field measurements and calibration with local stations</p> <p>III - Moderately robust models based on limited measurements and relying on remote stations from site</p> <p>IV - Models derived from remote stations, literature, and not calibrated</p>
1.9 Assessment of borrow resources (quantities and qualities) of construction materials over life of construction	<p>I - Comprehensive borrow search investigation program to verify availability and quality of construction materials, limited uncertainties or gaps</p> <p>II - Good investigation program, some uncertainties or gaps on quality and/or volumes</p> <p>III - Average investigation program, significant uncertainties and gaps exist on quality and/or volumes</p> <p>IV - Weak investigation program, poor understanding of quality and volume of resources</p>
Design – Laboratory Testing	
2.1 Laboratory test program development, level of collaboration to date	<p>I - Highly connected, adapted and updated as necessary to address gaps, with thorough collaboration and review</p> <p>II - Some linkage, some gaps still exist, with regular collaboration and review</p> <p>III - Limited linkage, limited collaboration and review</p> <p>IV - Limited linkage, non-collaborative and without review</p>
2.2 Testing and verification between lab results and field investigations	<p>I - Complete and results support analysis</p> <p>II - Good support to analysis and in situ measurements</p> <p>III - Supports primary analysis</p> <p>IV - Results limited, not validated</p>
2.3 Lab/Sampling QA/QC program	<p>I - Comprehensive program to verify quality/validity of results with limited uncertainties or gaps</p> <p>II - Good program, some uncertainties and gaps still exist</p> <p>III - Average program, significant uncertainties and gaps exist</p> <p>IV - Weak program</p>
2.4 Current understanding of geochemistry and physical properties of tailings (or other contained material)	<p>I - High level of certainty</p> <p>II - Good level of certainty, some gaps may exist</p> <p>III - Average level of certainty, clear gaps exist</p> <p>IV - Low level of certainty</p>
2.5 Current understanding of geochemistry and physical properties of materials used for construction	<p>I - High level of certainty</p> <p>II - Good level of certainty, some gaps may exist</p> <p>III - Average level of certainty, clear gaps exist</p> <p>IV - Low level of certainty</p>

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
2.6 Reporting, completeness, accessibility, and security of information	<p>I - Very complete reporting; present and future accessibility, very robust management/backup systems, information and history protected</p> <p>II - Good quality reporting; good to solid information retention, readily accessible and secure</p> <p>III - Average to good quality reporting; information moderately organized and retained, access may be difficult for some</p> <p>IV - Incomplete reporting, uncertainty in retention of information</p>
Design – Analysis and Documentation	
3.1 Site selection study (historical)	<p>I - Very detailed, thorough consultation and review</p> <p>II - Detailed, evident consultation and review</p> <p>III - Limited in scope, focus on compliance requirements</p> <p>IV - Pre-defined site, no consultation, no review</p>
3.2 Design Basis Memorandum (DBM) development, collaborative review through the project life to date and alignment with current state of practice	<p>I - Detailed, thoroughly reviewed, and meets current state of practice</p> <p>II - Good, evident review and updates at different project stages, does not meet current state of practice</p> <p>III - Document set at beginning of process, based on compliance requirements, limited review and changes</p> <p>IV - Limited information on design criteria</p>
3.3 Stability Analysis, based on current understanding from investigations, and failure modes	<p>I - Extensive, involving sensitivity analysis of multiple failure modes and conditions, using knowledge obtained from investigation, construction and operation phases, and thoroughly reviewed</p> <p>II - Detailed, involving many failure modes, good retroaction from investigation, construction and operation, evident review</p> <p>III - Fairly complete but focusing on meeting compliance requirements, limited retroaction from investigation, construction and operation, limited review</p> <p>IV - Limited in scope</p>
3.4 Flood routing (capacity and discharge, design storm events)	<p>I - System has the capacity for the design storm events, climate change considered, discharge mechanisms are more than adequate</p> <p>II - System has the capacity for the design storm events, discharge mechanisms are adequate</p> <p>III - System does not have the capacity for the design storm events or discharge mechanisms are inadequate</p> <p>IV - System does not have the capacity for the design storm events and discharge mechanisms are inadequate</p>
3.5 Hydrogeological analysis, based on current	<p>I - Deep understanding of hydrogeological conditions, analysis reviewed and currently valid, well calibrated on representative</p>

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
understanding from investigations and projections	field data, done to evaluate short- and long-term performance of facility II - Good understanding of hydrogeological conditions, evidence of review and analysis currently valid, limited calibration, done to evaluate short-term performance of facility III - Limited review, general understanding of flow conditions IV - Limited analysis
3.6 Risk analysis, based on current understanding from investigations, and failure modes	I - Detailed, up to date, to identify variety of risks and mitigations, strong focus on failure mode analysis and determination of possible consequences II - Completed, up to date, to identify key risks and mitigations, some failure modes analyses complete, consequences not fully understood III - Completed to identify key risks and mitigations, failure modes analysis incomplete IV - Conceptual level risk assessment
3.7 Review process through every design phase, including raises	I - Solid and formalized internal and external process, with appropriate reconciliations, for every design phase II - Internal and external process in place, with appropriate reconciliations, and working well for most design phases III - Review process limited in scope, not reconciled, or only for primary (or most recent) design phase IV - Weak or no review process
3.8 Reporting, completeness, accessibility, and security of information	I - Very complete reporting; present and future accessibility, very robust management/backup systems, information and history protected II - Good quality reporting; good to solid information retention, readily accessible and secure III - Average to good quality reporting; information moderately organized and retained, access may be difficult for some IV - Incomplete reporting, uncertainty in retention of information
Construction Management and Quality Assurance / Quality Control	
4.1 Project management program	I - In place and robust to allow work to proceed, through all construction phases II - In place and good coordination, most construction phases III - Limited, and satisfactory for only some construction phases IV - Limited program
4.2 Supervision program through life of construction	I - Dedicated qualified supervisory personnel on a full-time basis with clear roles and responsibilities, through all construction phases

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
	II - Supervisory personnel present on site in quasi-continuous basis, or satisfactory for most construction phases III - Supervisory personnel present on site periodically IV - Limited supervision on site
4.3 Construction QA/QC program through life of construction	I - Detailed, agreed, and followed QA/QC program, through all construction phases II - Solid QA/QC program being followed for most construction phases III - Minimum QA/QC program in place and/or evidence that construction is non-compliant with design IV - Limited QA/QC information collected, quality of construction unknown
4.4 Monitoring of infrastructure and foundation performance through life of construction	I - Monitoring as scheduled, and adapted construction rates to observed conditions through all construction phases II - Monitoring as scheduled, most construction phases III - Monitoring on occasional basis, or satisfactory only for some construction phases IV - Weak to non-existent monitoring
4.5 Construction reports through life of construction	I - Accessible, robust as-built reports and drawings, daily construction reports, including photos, etc., strong management/backup systems, to capture and maintain the history of the construction, for all construction phases II - Readily accessible and secure, good to solid reporting focused on as-built drawings and reports, considered satisfactory for most construction phases III - Reporting limited to site memos, or satisfactory for only some construction phases, moderately organized and retained, access may be difficult for some IV - Limited information, report and/or document control
4.6 Construction meetings between Owner and Contractor. Involvement of Designer (or rep) and Engineer of Record (as applicable) for critical aspects, through life of construction	I - Frequent meetings, ongoing communications II - Periodic meetings on site, or communications considered satisfactory for most construction phases III - Rare meetings, uncertainty of communication over life of construction IV - Limited knowledge of meetings
4.7 Review process through life of construction	I - Formal internal and external process, with appropriate reconciliations, through life of construction II - Internal and external process in place, with appropriate reconciliations, satisfactory for most construction phases III - Some internal or external process, limited reconciliations, in place IV - Limited to non-existent review

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
4.8 Management of change through life of construction	<p>I - Designer and Engineer of Record (EOR, as applicable) highly involved in the construction works and with all proposed changes, with scheduled site visits during the life of construction. Design changes accepted upon by all parties.</p> <p>II - Designer and EOR (as applicable) involved in construction works, changes approved, occasional site visits</p> <p>III - Designer and EOR (as applicable) involvement limited, notified of changes, no site visits</p> <p>IV - Designer and EOR (as applicable) not involved, uncertainty regarding knowledge of changes</p>
Operation and Monitoring	
5.1 Governance	<p>I - Owner with strong and clear governance according to current MAC guidelines (fully implemented, qualified personnel, clear roles and responsibilities)</p> <p>II - Owner with clarity on roles and responsibilities, MAC guideline not yet fully implemented</p> <p>III - Roles and responsibilities defined, but not fully assigned</p> <p>IV - Lack of clarity on roles and responsibilities</p>
5.2 Management system (e.g. TSM)	<p>I - Implemented, audited, results showing strong commitment</p> <p>II - Implemented and audited, improvements in progress</p> <p>III - In process of implementation</p> <p>IV - No formal system in place</p>
5.3 Operation, Maintenance and Surveillance Manual	<p>I - Up to date, annually reviewed and audited, and implemented</p> <p>II - Up to date, occasionally audited, and implemented</p> <p>III - Exists but not up to date, not fully followed</p> <p>IV - No clear OMS in place</p>
5.4 Monitoring Program and System	<p>I - In place, comprehensive coverage, with data acquisition system, proper real-time analysis and follow-up</p> <p>II - In place, comprehensive coverage, with follow-up on regular basis</p> <p>III - In place, partial coverage, and occasional follow-up</p> <p>IV - Limited monitoring and/or lack of follow-up</p>
5.5 Disclosure and External Consultation	<p>I - On-going disclosure, open and regular consultation with different stakeholders on performance of facility</p> <p>II - Good consultation with different stakeholders, disclosure beyond compliance requirements</p> <p>III - Limited consultation, disclosure in intent of meeting compliance requirements</p> <p>IV - Disclosure does not meet compliance</p>
5.6 Containment infilling / development planning (in	<p>I - Detailed, regularly updated, properly communicated and integrated into operations</p> <p>II - Updated annually, and communicated</p>

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
consideration and respective of DBM)	III - Infilling considered, strategy is mainly reactive with little flexibility IV - No formal infilling plan
5.7 Emergency preparedness program	I - In place and tested II - In place but not tested III - Exists but not up to date, nor tested IV - Not available
5.8 Formal review process (internal and external)	I - Operational, collaborative and open II - Operational, partially collaborative and open - only one of internal or external process formally exists III - Exists but no evidence of collaborative process IV - Weak or non-existent
5.9 Follow-up on all improvement recommendations and concerns	I - No recommendations for changes or all recommended changes have been implemented II - Recommended changes to address critical issues have been implemented, additional measures under consideration or being planned III - Recommended changes to address critical issues are in progress IV - No response to recommended changes
5.10 Management of Operational Risks	I - Risk registry up to date, fully communicated and implemented II - Risk registry recently updated, implementation in progress III - Risk registry created, not up to date IV - Risks assessed are limited in scope and general
Performance	
6.1 Geotechnical performance of infrastructure (movement, settlement, tension cracks, etc.)	I - Performance within design allowances II - Unexpected small to moderate changes, but stabilized, monitoring ongoing, and/or responsive actions complete to stabilize past or current performance issues III - Unexpected moderate to significant changes in performance, under close scrutiny, remedial actions in progress IV - Constant or increasing rates of movement or settlement, remedial plans under development
6.2 Hydrogeological performance of infrastructure (seepage, piezometric levels, internal erosion, impact to groundwater)	I - Performance within design allowances II - Unexpected small to moderate changes in performance, monitoring ongoing, and/or responsive actions complete to remediate past or current performance issues III - Unexpected moderate to significant changes in performance, or sign of internal erosion, under close scrutiny, remedial actions in progress

Table 2. List of Level of Practice criteria (cont'd)

Criteria	Level Descriptions
	IV - Constant or increasing rates of seepage or pore pressures and/or internal erosion, remedial plans under development
6.3 Hydrological performance of infrastructure (inclusive of freeboard, beach and spillway management and operations)	I - Performance according to the DBM at all times II - Minor temporary deviations of performance from DBM III - Frequent deviations of performance from DBM IV - Performance are not in agreement with DBM on a regular basis
6.4 Performance of foundation (movement and/or settlement)	I - Performance within design allowances II - Unexpected small to moderate changes, but stabilized, monitoring ongoing, and/or responsive actions complete to stabilize past or current performance issues III - Unexpected moderate to significant changes in performance, under close scrutiny, remedial actions in progress IV - Constant or increasing rates of movement or settlement, remedial plans under development

Draft

Table 3. Number of Level of Practice criteria and weighting for Silva et al. (2008), Julien et al. (2019) and the current work.

Performance categories	Design			Construction	Operation and Monitoring
	Investigation	Laboratory Testing	Analysis and Documentation		
Silva et al. (2008)					
Ponderation	20%	20%	20%	20%	20%
Number of criteria	7	4	6	3	3
Julien et al. (2019)					
Ponderation	20%	10%	20%	20%	30%
Number of criteria	9	6	7	8	13
Current work					
Ponderation	20%	13%	18%	18%	31%
Number of criteria	9	6	8	8	14

Table 4. Level of Practice for DCO Score

Level of Practice, LOP	DCO Score
I - High	1.0 to 1.25
Ib	> 1.25 to 1.75
II - Good	> 1.75 to 2.25
IIb	> 2.25 to 2.75
III- Average	> 2.75 to 3.25
IIIb	> 3.25 to 3.75
IV - Poor	> 3.75 to 4.0

Draft

Table 5. Probability ratings

Annual Probability of Failure	Probability Rating	Probability Rating Value
$\leq 1 \times 10^{-6}$	Negligible	0
$> 1 \times 10^{-6}$ to 1×10^{-5}	Very Low	1
$> 1 \times 10^{-5}$ to 1×10^{-4}	Low	2
$> 1 \times 10^{-4}$ to 1×10^{-3}	Moderate	3
$> 1 \times 10^{-3}$ to 1×10^{-2}	High	4
$> 1 \times 10^{-2}$	Very High	5

Draft

Table 6. Risk classification chart (adapted from Agnico Eagle Mines Ltd., 2020).

	Probability Rating (Value)					
	Negligible (0)	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Consequence Rating (Value)	Risk Rating (Value)					
Extreme/ Critical (5)	Negligible (0)	Medium (5)	Medium (10)	High (15)	Very High (20)	Very High (25)
Major (4)	Negligible (0)	Low (4)	Medium (8)	High (12)	High (16)	Very High (20)
Moderate (3)	Negligible (0)	Low (3)	Medium (6)	Medium (9)	High (12)	High (15)
Minor (2)	Negligible (0)	Low (2)	Low (4)	Medium (6)	Medium (8)	Medium (10)
Negligible (1)	Negligible (0)	Low (1)	Low (2)	Low (3)	Low (4)	Medium (5)

Table 7. Detailed results of risk assessment of the sample TSF portfolio.

Infrastructure	DCO Scores ¹						Ratings ²					
	D – I	D – LT	D – A&D	CM & QA/QC	O&M	Total	LOP	FS	APF	CR	PR	RR
TEX 1	0.31	0.21	0.26	0.23	0.39	1.40	I(b)	2.3	1.0E-08	4	0	0
TEX 2	0.29	0.21	0.26	0.23	0.37	1.36	I(b)	2.3	1.0E-08	4	0	0
TEX 3	0.42	0.30	0.33	0.33	0.50	1.88	II	1.6	2.2E-05	4	2	8
TEX 4	0.50	0.37	0.35	0.40	0.44	2.06	II	1.8	3.0E-06	4	1	4
TEX 5	0.26	0.16	0.23	0.23	0.40	1.28	I(b)	2.3	1.0E-08	4	0	0
TEX 6	0.26	0.16	0.21	0.19	0.44	1.26	I(b)	2.0	1.0E-08	4	0	0
TEX 7	0.29	0.19	0.23	0.21	0.37	1.29	I(b)	2.3	1.0E-08	4	0	0
TEX 8	0.26	0.19	0.23	0.21	0.37	1.26	I(b)	2.3	1.0E-08	4	0	0
TEX 9	0.24	0.16	0.16	0.21	0.33	1.10	I	2.5	1.0E-08	4	0	0
TEX 10	0.50	0.42	0.42	0.65	0.53	2.52	II(b)	1.7	1.5E-04	3	3	9
TEX 11	0.45	0.42	0.41	0.65	0.53	2.46	II(b)	1.9	9.5E-06	3	1	3
TEX 12	0.50	0.42	0.42	0.65	0.56	2.55	II(b)	2.2	1.0E-06	4	0	0
TEX 13	0.42	0.35	0.44	0.39	0.54	2.14	II	1.6	6.5E-05	5	2	10
TEX 14	0.42	0.35	0.44	0.39	0.60	2.21	II	1.5	3.2E-04	5	3	15
TEX 15	0.40	0.35	0.44	0.39	0.65	2.23	II	1.6	1.0E-04	5	2	10
TEX 16	0.42	0.35	0.44	0.39	0.54	2.14	II	2.6	1.0E-07	5	0	0
TEX 17	0.35	0.30	0.47	0.39	0.54	2.05	II	1.6	4.1E-05	5	2	10
TEX 18	0.37	0.30	0.47	0.39	0.54	2.07	II	1.7	1.2E-05	5	2	10
TEX 19	0.35	0.30	0.47	0.39	0.54	2.05	II	1.6	4.1E-05	5	2	10
TEX 20	0.44	0.33	0.44	0.39	0.54	2.14	II	1.4	8.7E-04	5	3	15
TEX 21	0.44	0.33	0.44	0.39	0.56	2.16	II	1.4	9.5E-04	5	3	15
TEX 22	0.44	0.33	0.44	0.39	0.58	2.19	II	1.4	1.0E-03	5	3	15
TEX 23	0.28	0.16	0.35	0.25	0.56	1.60	I(b)	1.6	8.5E-06	4	1	4
TEX 24	0.30	0.26	0.35	0.35	0.60	1.86	II	1.5	8.3E-05	4	2	8
TEX 25	0.28	0.14	0.35	0.21	0.49	1.47	I(b)	2.7	1.0E-08	4	0	0
TEX 26	0.28	0.14	0.35	0.21	0.51	1.49	I(b)	2.6	1.0E-08	4	0	0

Table 7. Detailed results of risk assessment of the sample TSF portfolio (cont'd)

Infrastructure	DCO Scores ¹						LOP	FS	APF	Ratings ²		
	D – I	D – LT	D – A&D	CM & QA/QC	O&M	Total				CR	PR	RR
TEX 27	0.28	0.14	0.35	0.33	0.53	1.63	I(b)	1.4	1.6E-04	4	3	12
TEX 28	0.35	0.21	0.33	0.19	0.58	1.65	I(b)	1.5	4.1E-05	4	2	8
TEX 29	0.60	0.44	0.49	0.53	0.56	2.63	II(b)	1.5	1.9E-03	4	4	16
TEX 30	0.33	0.16	0.26	0.26	0.53	1.54	I(b)	1.9	6.5E-08	4	0	0
TEX 31	0.33	0.30	0.33	0.35	0.66	1.97	II	1.6	2.9E-05	2	2	4
TEX 32	0.21	0.14	0.23	0.21	0.40	1.19	I	1.3	1.7E-04	1	3	3
TEX 33	0.35	0.26	0.39	0.21	0.58	1.79	II	1.9	2.4E-07	5	0	0
TEX 34	0.35	0.26	0.39	0.21	0.60	1.81	II	1.9	2.6E-07	5	0	0
TEX 35	0.35	0.26	0.39	0.21	0.60	1.81	II	2.2	1.0E-08	5	0	0
TEX 36	0.35	0.26	0.39	0.21	0.58	1.79	II	1.5	6.6E-05	5	2	10
TEX 37	0.46	0.35	0.42	0.56	0.47	2.26	II(b)	2.8	1.0E-07	4	0	0
TEX 38	0.46	0.35	0.42	0.56	0.56	2.36	II(b)	1.4	1.9E-03	4	4	16
TEX 39	0.46	0.35	0.42	0.56	0.56	2.36	II(b)	1.6	2.7E-04	4	3	12

¹ D – I: Design – Investigation, D – LT : Design – Laboratory testing, D – A&D: Design – Analysis & documentation, CM & QA/QC: Construction management & quality-assurance/quality-control, O&M: Operation & monitoring, P: Performance

² CR: Consequence rating, PR: probability rating, RR: risk rating

Figure 1. Empirical relationships for annual probability of failure vs factor of safety for different levels of engineering (adapted from Silva et al., 2008).

119x139mm (96 x 96 DPI)

Figure 2. Comparison of annual probability of failure – factor of safety relationships from the Silva Method and the proposed method.

229x137mm (96 x 96 DPI)

Figure 3: Example risk classifications chart.

219x119mm (96 x 96 DPI)

Figure 4. DCO scores and level of practice of sample TSF portfolio.

273x170mm (96 x 96 DPI)

Figure 5. Estimated annual probabilities of failure of sample TSF portfolio.

219x136mm (96 x 96 DPI)

Figure 6. Risk classification chart with results of sample TSF portfolio.

219x119mm (96 x 96 DPI)

Figure 7. Plan view of the Mount Polley tailings storage facility showing location of cross-sections evaluated for the case study (adapted from BCMEM, 2015).

241x172mm (150 x 150 DPI)

Figure 8. Results of risk assessment of the Mount Polley tailings storage facility: a) annual probability of failure; and b) risk classification.

246x98mm (96 x 96 DPI)