*Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)*
*Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5*
*https://doi.org/10.53243/ICSMGE2026-2101*

# Reliability analysis based on the response surface method applied to the static liquefaction failure of a tailings dam

**Giosser Arica**, Gregory Alvarado, Johann Suarez, Fabiola Serazo
*Consultant, SRK Consulting, Lima, Peru, garica@srk.com.pe, galvarado@srk.com.pe, jsuarez@srk.com.pe, fserazo@srk.com.pe*

ABSTRACT: Recent failures of upstream-raised tailings storage facilities (TSF) have raised significant concerns in recent years. The problem is mainly based on the dependence of the stability of the structure on the strength of the tailings, since most of them have been deposited in a loose, saturated, and normally consolidated state. Therefore, it is necessary to consider the uncertainty of the tailings parameters when evaluating the stability conditions by deterministic methods. In this paper, the exercise of evaluating physical stability and performing a reliability analysis of an upstream tailings dam using the response surface method is developed. Three types of response functions are employed for this analysis: The quadratic polynomial without cross terms (the most popular function), the quadratic polynomial with cross terms, and the Support Vector Machine (SVM), a machine learning-based technique. These functions enable the estimation of the Load Factor (LF) by applying a gravity load multiplier to the system through simulations conducted in a finite element software (Plaxis 2D), utilizing advanced constitutive models such as NorSand and Hardening Soil with Small Strain (HSS). Furthermore, this analysis accounts for the variability of the peak and residual undrained shear strength of liquefiable, contractive materials, complemented by the effect of the variability in the phreatic surface elevation. Finally, explicit solutions obtained from the response surface methodology are used for reliability analysis. The methods employed for this analysis include the First Order Reliability Method (FORM) and the Second Order Reliability Method (SORM). This approach enables the comparison of results – specifically, failure probability and reliability index – obtained from the three response surface functions and the two reliability methods mentioned above.

KEYWORDS: Static Liquefaction, Response Surface Method, Tailings Dam, Reliability Analysis, Support Vector Machine (SVM).

## 1 INTRODUCTION

Upstream tailings storage facilities (TSF) are widely used due to their low cost but are inherently vulnerable to failure. Past disasters, such as Merriespruit (1994), Fundao (2015), and Brumadinho (2019), revealed static liquefaction as the dominant failure mechanism, with catastrophic consequences. This phenomenon is linked to the contractive behavior of fine tailings, which are typically deposited in a loose, saturated, and normally consolidated state (Jefferies & Been, 2015).

Conventional deterministic analyses, based solely on the Factor of Safety (FoS), overlook the inherent variability of soil parameters and may result in unconservative designs (Phoon & Kulhawy, 1999). To address this gap, probabilistic approaches have been developed, combining advanced numerical simulations with surrogate modeling techniques to evaluate the uncertainty in geotechnical parameters and their influence on stability (Kar & Roy, 2022; Nagula et al., 2023).

This study presents an integrated framework for the reliability analysis of an upstream TSF subject to static liquefaction. The methodology combines advanced constitutive models (NorSand and HSS) used in deformation analysis, considering a gradual increase of gravity loading as the triggering mechanism of static liquefaction, surrogate functions via Response Surface Method (RSM), and reliability assessment using FORM and SORM. The goal is to identify the most influential variables and quantify the probability of failure, thereby supporting safer TSF design and management.

## 2 CASE STUDY

The analyzed TSF was constructed using the upstream method, in which cyclone coarse tailings (cyclone sand) were placed along the perimeter to form the containment dikes, thereby retaining the fine tailings (slimes) within the impoundment. The facility has been in operation since 2000 and has undergone four rises of approximately 8 m each, reaching a total height of 35 m while maintaining a 1 m freeboard at the crest.

Hydrogeological characterization, based on piezometer data and numerical modeling, defined the phreatic surface configuration and scenarios adopted in the analysis. A complementary hydrological assessment supported treating the phreatic level as a random variable in the reliability framework.

The TSF is instrumented with inclinometers, piezometers, accelerometers, and remote InSAR monitoring, all of which indicate stable behavior, with no significant displacements or abnormal pore pressure trends. This validates the use of calibrated numerical models under current conditions to project hypothetical critical scenarios. More than two decades of geotechnical investigations, including test pits, boreholes, SPT, LPT, SCPTu, MASW, and special laboratory tests (primarily triaxial and direct simple shear tests), have enabled the definition of representative geotechnical units for the dam and its foundation. Geotechnical parameters were characterized for both drained and undrained conditions, and the variability of shear strength was explicitly incorporated within a probabilistic framework.

### 2.1 *Geotechnical characterization*

In the evaluated tailings storage facilities, five (05) geotechnical units were identified, defined based on the geological information available for the study. The physical characteristics of each unit are presented in the following table.

Table 1. Average physical characteristics of the geotechnical units.

| Geotechnical units | USCS | Fines Con tent [%] | Water Content [%] |
|---|---|---|---|
| UG1: Coarse tailings | SM | 20 | 20 |
| UG2: Fine tailings | ML | 80 | 25 |
| UG3: Fine foundation | CL | 65 | 35 |
| UG4: Gravelly foundation | GM | 34 | 20 |
| UG5: Bedrock | - | - | - |

Before estimating the geotechnical parameters, a liquefaction susceptibility assessment was performed using the methods proposed by Fear & Robertson (1995), Olson (2009), and Robertson (2016). The results indicate that the coarse tailings

(UG1), fine tailings (UG2), and fine foundation (UG3) exhibit a high susceptibility to flow liquefaction.

In addition, pore pressure ratios ($B_q$) were derived from SCPTu tests and grouped by geotechnical unit. UG1 and UG2 show $B_q > 0.0$ in several zones, while UG3 reaches $B_q > 0.5$, confirming that these materials behave under undrained conditions below the phreatic surface. According to Monforte (2022), they are considered fully saturated in this study.

Fine tailings (UG2) and fine foundation soils (UG3) exhibit softening and contractive behavior, as indicated by results from triaxial, DSS, SCPTu, and field VST tests. To properly capture this contractive response and the susceptibility of certain materials to flow liquefaction, the Hardening Soil with Small Strain (HSS) and NorSand constitutive models were calibrated against the tests. However, both units showed significant variability in their normalized undrained shear strength values at both peak and residual ratios.

For the fine tailings (UG2), peak undrained shear strength ratios $USR = s_u/\sigma'_{v0}$ was estimated from vane shear tests, yielding values between $0.17 < USR < 0.30$. Complementary, SCPTu data, filtered for $B_q > 0.02$ and $B_q > 0.20$ (indicating excess pore pressure), produced a 50% exceedance value of 0.33. Consolidated undrained triaxial tests with strain-softening response confirmed the contractive tendency, with peak strength ratios $0.24 < USR < 0.29$. From probability density and cumulative distribution functions, a mean peak strength ratio $USR = 0.285$ and a range $0.25 < USR < 0.32$ were adopted. For residual strength, integrating SCPTu, vane, and monotonic post-cyclic laboratory tests yielded a mean value $USR_{res} = 0.16$ and a range $0.12 < USR_{res} < 0.20$.

For the fine foundation (UG3), SCPTu tests filtered for $B_q > 0.02$ and $B_q > 0.50$ defined a mean peak strength ratio $USR = 0.265$, within a range of $0.21 < USR < 0.32$. Residual strength, derived from combined SCPTu, vane, and laboratory data, resulted in a mean $USR_{res} = 0.14$, and a range $0.10 < USR_{res} < 0.48$. These values, derived from an integrated field and laboratory testing program, were modeled using log-normal and normal distributions, thereby capturing the inherent variability of the materials and providing the statistical basis for the reliability analysis.

## 2.2 *Cases analyzed*

Base case adopted parameters are presented in Table 2. Fine tailings and fine foundation soils exhibit considerable variability in their normalized peak and residual undrained shear strength ratios. To capture this behavior within the NorSand constitutive model, two parameters were varied. The variability of the peak undrained shear strength ratio $USR$ was reproduced by adjusting the hardening parameter ($H_\psi$), whereas the variability of the residual undrained shear strength ratio $USR_{res}$ was represented through modifications of the state parameter ($\psi'_0$) (Table 3). Figures 1 and 2 illustrate the simulations of the undrained shear strength ratio for the minimum and maximum values of the state parameter and hardening parameter in both geotechnical units.

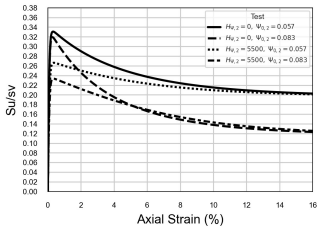


Figure 1. NorSand-based simulation of undrained shear strength ratio – UG2 (fine tailings).

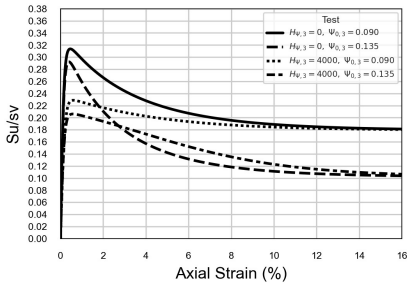


Figure 2. NorSand-based simulation of undrained shear strength ratio – UG3 (Fine foundation).

Table 2. Mean material parameters.

| Parameter | Symbol | UG1 | UG2 | UG3 | UG4 | UG5 |
|---|---|---|---|---|---|---|
| Constitutive model | - | HSS NS | HSS NS | HSS NS | HSS | LE |
| Unit weight (kN/m$^3$) | $\gamma$ | 19 | 17 | 15 | 21 | - |
| Friction angle (°) | $\phi'$ | - | - | - | 36 | - |
| Effective cohesion (kPa) | $c'$ | - | - | - | 1 | - |
| Threshold shear strain (-) | $\gamma_{0.7}$ | - | - | - | 1.0 E-04 | - |
| Reference shear modulus (MPa) | $G_0^{ref}$ | - | - | - | 240 | - |
| Unloading/reloading stiffness (MPa) | $E_{ur}^{ref}$ | - | - | - | 120 | - |
| Secant stiffness (MPa) | $E_{50}^{ref}$ | - | - | - | 40 | - |
| Tangent stiffness (MPa) | $E_{oed}^{ref}$ | - | - | - | 45 | - |
| Elastic stiffness (GPa) | $E$ | - | - | - | - | 3.5 |
| Exponent of power law (-) | $m$ | - | - | - | 0.5 | - |
| Poisson's ratio (-) | $\nu$ | 0.2 | 0.2 | 0.3 | 0.2 | 0.3 |
| Failure ratio (-) | $R_f$ | - | - | - | 0.9 | - |
| At rest pressure coefficient (-) | $K_0^{nc}$ | - | - | - | 0.5 | - |
| Exponent of the power-law elasticity | $n_G$ | 0.5 | 0.5 | 0.5 | - | - |
| Friction ratio at critical state (-) | $M_{tc}$ | 1.46 | 1.45 | 1.30 | - | - |
| Dilation parameter (-) | $N$ | 0.05 | 0.05 | 0 | - | - |
| Slope of min dilatancy vs $\psi$ (-) | $X_{tc}$ | 2.0 | 2.0 | 2.0 | - | - |
| Hardening parameter (-) | $H_0$ | 300 | 330 | 280 | - | - |
| Hardening parameter (-) | $H_\psi$ | 3000 | 3000 | 2000 | - | - |
| Initial state parameter (-) | $\psi_0$ | 0.03 | 0.07 | 0.11 | - | - |
| Void ratio at 1 kPa | $\Gamma$ | 0.81 | 0.80 | 1.65 | - | - |
| Slope of the CSL (-) | $\lambda_e$ | 0.065 | 0.05 | 0.08 | - | - |

In addition, the variability of the phreatic surface in the tailings storage facility was considered, considering the permeability of the fine tailings and the fluctuations in piezometric levels recorded during both dry and wet seasons. The phreatic surface is estimated to fluctuate by up to 5 m.

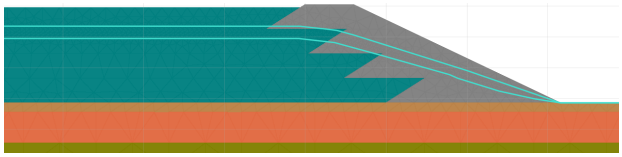

Figure 3. Phreatic level variation across the evaluated section.

Table 3. Value options for each variable parameter.

| Index | $H_{\psi,2}$ | $H_{\psi,3}$ | $\psi_{0,2}$ | $\psi_{0,3}$ | WT* |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.0570 | 0.0900 | 0 |
| 1 | 1500 | 1000 | 0.0635 | 0.1000 | 1 |
| 2 | 3000 | 2000 | 0.0700 | 0.1100 | 2 |
| 3 | 4250 | 3000 | 0.0765 | 0.1225 | 3 |
| 4 | 5500 | 4000 | 0.0830 | 0.1350 | 4 |

(*) phreatic surface index: 0 lowest to 4 highest water table.

## 3 NUMERICAL MODELING

The model geometry incorporated the five phreatic surface scenarios previously defined. The mesh consisted of 3,369 fifteen-node triangular elements (27,239 nodes in total), with finer discretization in the zone affected by phreatic variability. Mesh quality verification yielded a minimum value of 0.2859, which is within the acceptable range for Plaxis models.

To reproduce the current stress state of the TSF, staged construction was simulated as follows:

- Initial Phase: initial stresses generated under $K_0$ conditions, with horizontal layers in equilibrium.
- Phases 1–3: successive raises of the dam, all materials modeled under drained conditions.
- Phase 4: final raise, including the phreatic surface configuration derived from piezometric data and its variability; materials in drained conditions.
- Phase 5: materials located below the phreatic level switched to undrained conditions.

After these stages were completed, gravity loads were increased while keeping the undrained shear strength constant. This load increase is effectively equivalent to a Load Factor in the LRFD method. The procedure, however, is not a full LRFD design method, as the strength was adopted as a full probability density function instead of a design value. In other words, no resistance factor was employed in the study. Instead, 70 scenarios were considered, varying the geotechnical parameters and the previously defined phreatic surface levels. A Python script was developed to automatically generate these models.

## 4 RESPONSE SURFACE METHOD

### 4.1 *Theorical Background*

The Response Surface Methodology (RSM) is a statistical tool designed to approximate complex models through explicit functions fitted from a limited number of simulations or experiments. The objective is to construct an approximated function $\hat{g}(X)$ that reproduces, with acceptable accuracy, the behavior of the original model $g(X)$ within a given analysis domain $D$:

$$\hat{g}(X) \approx g(X)\,,\ X \in D \tag{1}$$

The approximation is generally obtained by multivariable regression based on a linear system:

$$A \bullet \theta = y,\ \theta = (A^T A)^{-1} A^T y \tag{2}$$

In this study, three formulations of RSM were considered. The first one is quadratic polynomials without cross terms (M1), which are frequently adopted due to their simplicity and computational efficiency. This model can represent both linear and quadratic effects of the independent variables, but does not capture possible interactions between them. The general form is given by:

$$\hat{g}(X) = a + \sum_{j=1}^{n} b_j X_j + \sum_{j=1}^{n} c_j X_j^2 \tag{3}$$

where $a$ denotes the intercept, $b_j$ the coefficients of the linear contributions of each variable $X_j$, $c_j$ the coefficients of the quadratic terms, $X_j$ represent the independent input variables, and $n$ the number of variables.

The second method is quadratic polynomials with cross terms (M2), which can provide a more versatile representation when interactions between variables become significant. By incorporating terms of the type $X_i X_j$, these models can reproduce coupled effects and asymmetric curvatures that the simpler quadratic form cannot. The general structure is:

$$\hat{g}(X) = a + \sum_{j=1}^{n} b_j X_j + \sum_{j=1}^{n} c_j X_j^2 + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} d_{ij} X_i X_j \tag{4}$$

Here, in addition to the parameters already defined, the coefficients $d_{ij}$ quantify the strength of the interaction between pairs of variables $X_i$ and $X_j$.

Finally, Support Vector Machines with a radial basis function (RBF) kernel (SVM) were also considered as a surrogate modeling technique within the RSM framework. Originally developed for classification, SVM has been successfully adapted for regression, offering the ability to capture nonlinear and multidimensional relationships through kernel functions. Their regression form can be expressed as:

$$\hat{g}(X) = \sum_{i=1}^{N} \left(\alpha_i - \alpha_i^*\right) K(X, X_i) + b \tag{5}$$

In this formulation, $\alpha_i$ and $\alpha_i^*$ are the Lagrange multipliers obtained during training, $K(X, X_i)$ represents the kernel function that maps the variables into a higher-dimensional space, and $b$ is the bias term.

Overall, polynomial functions are suitable for problems that require computational efficiency and moderate accuracy, while SVM provides a robust alternative for complex, nonlinear systems.

### 4.2 *Application to Tailings Dam Stability*

In assessing tailings dam stability, traditional deterministic approaches offer a limited perspective. These methods rely on mean values of soil parameters and do not explicitly account for their inherent variability which may lead to either underestimating or overestimating the level of safety. To overcome these limitations, this study applies the Response Surface Methodology (RSM) as a surrogate modeling tool to approximate the relationship between key geotechnical variables and the Load Factor (LF), enabling its integration into reliability analyses.

An “equivalent FoS” is defined as the ratio of the capacity (as a statistical variable) and the load factor (represented by the increment of gravity loading in each case). This “equivalent FoS” is of probabilistic nature and should not be confused with the classical concept of factor of safety which is deterministic in nature. The equivalent FoS was represented by functions fitted from numerical simulations performed in Plaxis, considering systematic variations of material properties and hydraulic conditions. The resulting models are simplified yet statistically representative, enabling the efficient evaluation of dam behavior under various parameter combinations.

The iterative construction of M1 and M2 refined the response surface in the vicinity of the critical design point, while the SVM provided greater flexibility in capturing nonlinear trends and complex variable interactions.

The initial results obtained with M1 (Figure 4) showed that the model reproduced exactly the FoS values from the initial experimental design, although its predictive capability was limited to the sampled cases. A second iteration (Figure 5) improved the local representation around the design point, highlighting the strong adverse influence of the water table (WT) and the sensitivity of the FoS to reductions in undrained shear strength. Nevertheless, M1 remained limited in its ability to capture coupled effects between variables.

The inclusion of cross terms in M2 provided a more versatile representation (Figure 6). This formulation not only reproduced the individual influence of each parameter but also quantified their combined effects, such as the interaction between WT and soil strength parameters. The second iteration (Figure 7) further refined the approximation near the critical state, improving both predictive accuracy and result consistency.

Finally, the SVM model trained with 70 simulation cases demonstrated superior predictive performance compared to the polynomial formulations. The use of the RBF kernel enabled the SVM to implicitly capture nonlinearities and complex interactions, resulting in highly accurate approximations with excellent parity between observed and predicted FoS values (Figure 8). Moreover, the SVM proved particularly effective in reproducing the combined effect of WT and undrained strength parameters, which are critical in determining stability. A comparative summary is presented in Table 4.

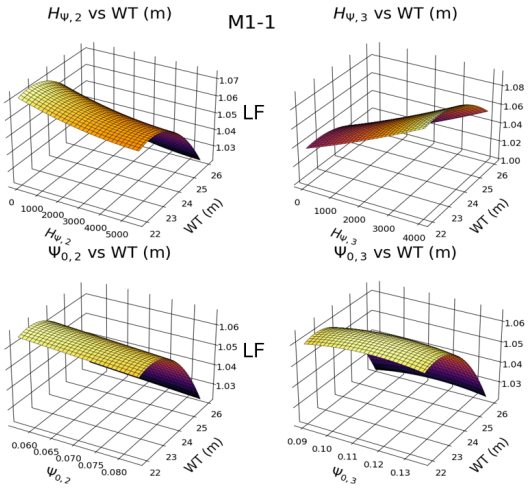


Figure 4. Response surface generated using Method M1 for the first iteration.

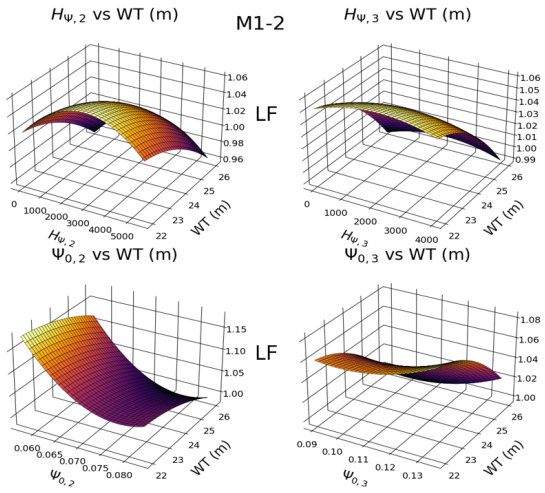


Figure 5. Response surface generated using Method M1 for the second iteration.

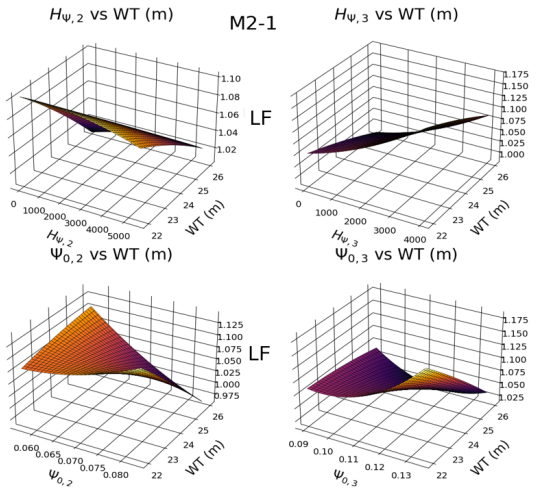


Figure 6. Response surface generated using Method M2 for the first iteration.

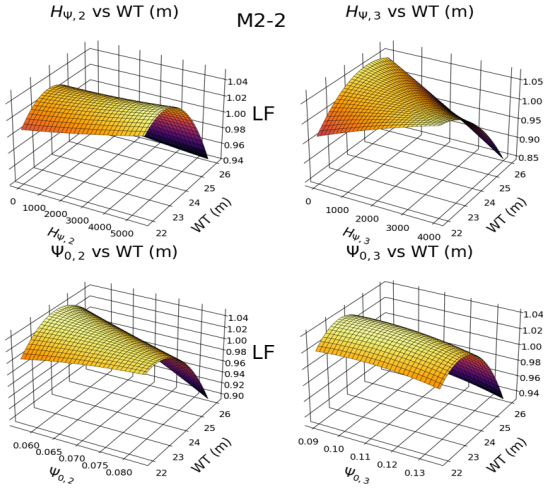


Figure 7. Response surface generated using Method M2 for the second iteration.

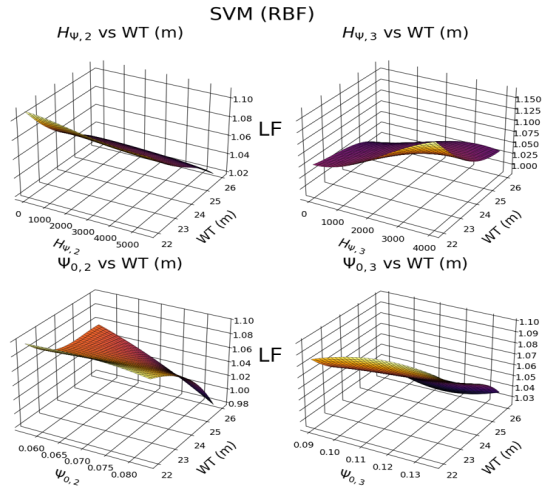


Figure 8. Response surface generated using the SVM method.

Overall, M1 offered a basic yet interpretable model suitable for preliminary analysis, M2 provided a better balance between complexity and accuracy, and the SVM achieved the highest predictive capacity, particularly in regions where nonlinear behavior dominates. This comparison highlights the importance of integrating polynomial response surfaces with advanced machine learning models to enhance the reliability assessments.

Table 4. Value options for each variable parameter.

| Criterion | M1 | M2 | SVM |
|---|---|---|---|
| Structure | 11 cases; fits basic curvature near design point | 21 cases; includes cross-terms | 70 cases; cross-validation and strict hyperparameter training |
| Performance | Weak, underestimates FoS in extremes | Better than M1 | Best within design domain |
| Parity | High scatter at low FoS | Less scatter than M1 | Best parity |
| WT Sensitivity | Limited; misses combined effects and cross-terms | Captures WT and strength variability better than M1 | Accurately captures nonlinear WT–strength interaction |
| Extrapolation | Poor beyond design point | Balanced fit, but limited to design range | Robust within training domain |
| Interpretability | High, simple coefficients | High, more coefficients | Medium, no explicit function |

## 5 RELIABILITY ANALYSIS

Reliability analysis provides a probabilistic framework to quantify the likelihood of failure by explicitly incorporating uncertainties in geomechanical parameters, boundary conditions, and loadings. This approach not only evaluates whether a system is safe but also estimates the probability that it may fail, thereby supporting more robust risk-informed decision-making. The reliability framework is built around the limit state function, $g(X)$, which separates safe from failure conditions:

$$g(X) = R(X) - S(X) \tag{6}$$

where $X = [X_1, X_2, \dots, X_n]$ is the vector of random variables, $R(X)$ is the system resistance, and $S(X)$ is the applied load. Failure occurs when $g(X) \leq 0$. The probability of failure ($PoF$) is thus defined as:

$$PoF = P[g(X) \leq 0] \tag{7}$$

To facilitate computations, the variables are commonly transformed into an equivalent standard normal space using Nataf or Rosenblatt transformations. Within this space, the reliability index $\beta$ is defined as the shortest distance from the origin to the limit state surface, serving as a measure of system safety. Its relationship with PoF is expressed as:

$$PoF = \Phi(-\beta) \tag{8}$$

where $\Phi$ denotes the standard normal cumulative distribution. Larger values of $\beta$ indicate lower failure probabilities.

### 5.1 *Methods of Reliability Analysis*

Several methods are available to estimate the PoF, depending on the system's complexity and the desired accuracy.

Monte Carlo Simulation (MCS): Based on repeated random sampling of the input variables, MCS provides a robust and conceptually simple estimate of $PoF$. Its main drawback is the high computational demand, particularly when failure probabilities are small. Variance reduction techniques such as importance sampling and stratified sampling are often employed to increase efficiency.

First-Order Reliability Method (FORM): FORM approximates the limit state surface locally by a tangent plane at the design point (the most probable point of failure in the standard normal space). The reliability index is computed as:

$$\beta = \|U^*\| = \sqrt{\sum_{i=1}^{n} \left(U_i^*\right)^2} \tag{9}$$

where $U^*$ is the design point. FORM is efficient and provides both $PoF$ estimates and sensitivity factors, identifying the contribution of each variable to system reliability.

Second-Order Reliability Method (SORM): Extends FORM by including the local curvature of the limit state surface through a second-order Taylor expansion. This correction improves accuracy when nonlinearities are significant. SORM estimates typically reduce $PoF$ compared to FORM, reflecting the curved nature of realistic failure surfaces.

### 5.2 *Application to the Tailings Dam*

The response surfaces developed in Section 4 (M1: quadratic polynomial without interactions, M2: quadratic polynomial with interactions, and SVM with RBF kernel) were employed as limit state functions. Reliability indices and failure probabilities were computed using both FORM and SORM.

The first iteration with the M1 model yielded $\beta \cong 5.19$ and $PoF \cong 0.00$, reflecting an over-optimistic estimate due to the limited design of experiments (DOE). The second iteration, however, reduced to $\beta \cong 1.76$ and increased $PoF \cong 4.0\%$, providing a more realistic measure of stability. For the M2 model, the first iteration produced $\beta \cong 2.22$ with $PoF \cong 0.44\%$. The refinement in the second iteration led to $\beta \cong 2.02$ and $PoF \cong 2.16\%$, reflecting an improved representation of the critical region while maintaining relatively high reliability.

Finally, using 70 cases, the SVM model delivered $\beta \cong 1.86$and a $PoF \cong 3.16\%$. These values are comparable to those from the refined M1 and M2 surfaces but demonstrate an enhanced ability to capture nonlinear interactions. The SORM estimates confirmed the same overall trends, with slightly lower $PoF$ values due to curvature adjustments. For example, the $PoF$ for M1-2 dropped from 3.96% (FORM) to 2.68% (SORM), while for M2-2 it decreased from 2.16% to 1.63%. The SVM also showed a reduction from 3.16% to 2.24%.

The reliability indices obtained in this study ($\beta \cong 1.76 - 2.02$) fall within the range typically reported for geotechnical systems like upstream-raised TSFs. For example, Liu et al. (2017) documented $\beta = 0.91$ for an unstable rock slope and $\beta = 3.03$ for a stable multi-layer slope. Accordingly, the evaluated TSF exhibits an intermediate and consistent level of stability, aligned with values observed in comparable applications.

Early iterations (e.g., M1-1) yielded unrealistic $\beta$ values and near-zero failure probabilities, highlighting the need to refine the DOE towards the critical region. Incorporating interaction terms in M2 improved model robustness and generally produced lower failure probabilities compared to M1. The SORM approach provided moderate corrections relative to FORM in the presence of nonlinearities, without altering the performance ranking among models. Overall, M2 emerged as a balanced surrogate, combining functional transparency with reasonable accuracy, while the SVM offered a powerful alternative for capturing strongly nonlinear responses.

## 6 CONCLUSIONS

This study applied reliability analysis based on Response Surface Methodology (RSM) to evaluate the static liquefaction of an upstream tailings dam. The integration of field and laboratory data enabled realistic calibration of constitutive models, capturing both contractive and strain-softening behavior of liquefiable materials. The combined use of NorSand and HSS constitutive models enabled an adequate representation of undrained strength and deformability, thereby reducing uncertainty in the estimation of the equivalent $FoS$.

Among the surrogate models, the quadratic polynomial with cross terms (M2) provided the best balance between

interpretability and accuracy, while the SVM achieved the lowest errors by capturing nonlinearities, albeit with reduced physical traceability. Reliability analyses with FORM and SORM confirmed consistent trends, with SORM slightly reducing failure probabilities due to curvature effects.

The water table was identified as the dominant controlling factor, systematically reducing stability and amplifying the effect of low undrained strengths. Numerical modeling further confirmed the importance of drainage conditions, mesh refinement, and phase definition in reproducing pore pressures and shear deformations.

Comparative assessment revealed that M1 provided a basic yet biased approximation, M2 corrected interaction effects while maintaining interpretability, and the SVM offered superior predictive performance in nonlinear domains. Reliability indices indicated that initial iterations tended to overestimate safety, while refined sampling stabilized results and provided more realistic values of the probability of failure.

Overall, the integrated framework—comprising numerical modeling, surrogate modeling, and reliability analysis—proved coherent, reproducible, and transferable, provided that the geotechnical characterization and design domains are properly defined and extrapolation is avoided.

The use of RSM and SVM provides clear practical benefits. Surrogate models substantially reduce the need for full numerical simulations, resulting in meaningful computational savings. Their rapid evaluation enables efficient exploration of scenarios and key parameters such as the water table, supporting faster and better-informed decision-making and strengthening the safety management of tailings facilities.

## 7 ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the Geoengineering Team of SRK Consulting (Peru) for their continuous support throughout the development of this research. Their collaboration, technical discussions, and the provision of valuable data and insights were essential to the successful completion of this study.

## 8 REFERENCES

Australian National Committee on Large Dams (ANCOLD). (2019). Guidelines on tailings dams: Planning, design, construction, operation and closure (Addendum, July 2019).

Been, K., Jefferies, M. G., & Hachey, J. (1991). The critical state of sands. Géotechnique, 41(3), 365–381. https://doi.org/10.1680/geot.1991.41.3.365

Bentley Systems. (2020). User defined soil models: NorSand—An elasto-plastic model for soil behaviour with static liquefaction (PLAXIS CONNECT Edition V21.00)

Castonguay, V., & Konrad, J.-M. (2019). A modified NorSand model for the prediction of static and cyclic behaviour of sands under simple shear loading. In F. Silvestri & N. Moraci (Eds.), Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions: Proceedings of the 7th International Conference on Earthquake Geotechnical Engineering (ICEGE 2019), (pp. 1647–1654). CRC Press. https://doi.org/10.1201/9780429031274-149

Hidalgo, C., & Assis, A. P. (2011). Evaluación de la incertidumbre en el análisis de estabilidad de un talud excavado en suelos residuales. En Actas del PanAm CGS 2011: 14th Pan-American Conference on Soil Mechanics and Geotechnical Engineering & 64th Canadian Geotechnical Conference. https://www.issmge.org/publications/publication/evaluacin-de-la-incertidumbre-en-el-anlisis-de-estabilidad-de-un-talud-excavado-en-suelos-residuales

Hidalgo Montoya, C. A., & Pacheco de Assis, A. (2011). Herramientas para análisis por confiabilidad en geotecnia: Aplicación. Revista Ingenierías Universidad de Medellín, 10(18), 79–86. https://revistas.udem.edu.co/index.php/ingenierias/article/view/339

Jefferies, M., & Been, K. (2015). Soil liquefaction: A critical state approach (2nd ed.). CRC Press. https://doi.org/10.1201/b19114

Jefferies, M. G., & Shuttle, D. A. (2011). On the operating critical friction ratio in general stress states. Géotechnique, 61(8), 709–713. https://doi.org/10.1680/geot.9.T.032

Kar, S. S., & Roy, L. B. (2022). Probabilistic Based Reliability Slope Stability Analysis Using FOSM, FORM, and MCS. Engineering, Technology & Applied Science Research, 12(2), 8236–8240. https://doi.org/10.48084/etasr.4689

Liu, X., Ni, W., Huang, L., Liu, Q., & Ma, J. (2017). Reliability Analysis of Tailings Dams: A Case Study in Jiangxi Province, China. Geo-Risk 2017, 178–187. https://doi.org/10.1061/9780784480724.017

Nagula, S. S., Liu, H., Nadim, F., Jostad, H. P., & Piciullo, L. (2023). Response surface based probabilistic studies on static liquefaction failure of tailings dams. En L. Zdravkovic, S. Kontoe, D. M. G. Taborda, & A. Tsiampousi (Eds.), Proceedings 10th NUMGE 2023: 10th European Conference on Numerical Methods in Geotechnical Engineering (pp. 1–6). International Society for Soil Mechanics and Geotechnical Engineering. https://doi.org/10.53243/NUMGE2023-201

Phoon, K.-K., & Kulhawy, F. H. (1999). Characterization of geotechnical variability. Canadian Geotechnical Journal, 36(4), 612–624. https://doi.org/10.1139/t99-038

Razavi, S. K., & Yniesta, S. (2023, October). On the numerical issues of the NorSand model in the simulation of undrained behavior of granular soils. Paper presented at the GeoSaskatoon 2023: 76th Canadian Geotechnical Conference, Saskatoon, SK, Canada.