We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Practical Framework for Probabilistic Analysis of Embankment Dams

Xiangfeng Guo and Daniel Dias

Abstract

Uncertainties, such as soil parameters variability, are often encountered in embankment dams. Probabilistic analyses can rationally account for these uncertainties and further provide complementary information (e.g., failure probability and mean/variance of a model response) than deterministic analyses. This chapter introduces a practical framework, based on surrogate modeling, for efficiently performing probabilistic analyses. An active learning process is used in the surrogate model construction. Two assessment stages are included in this framework by respectively using random variables (RV) and random fields (RF) for the soil variability modeling. In the first stage, a surrogate model is coupled with three probabilistic methods in the RV context for the purpose of providing a variety of useful results with an acceptable computational effort. Then, the soil spatial variability is considered by introducing RFs in the second stage that enables a further verification on the structure reliability. The introduced framework is applied to an embankment dam stability problem. The obtained results are validated by a comparison with direct Monte Carlo Simulations, which also allows to highlight the efficiency of the employed methods.

Keywords: embankment dam, slope stability, reliability analysis, sensitivity analysis, random field

1. Introduction

According to the International Commission of Large Dams (ICOLD) database updated in September 2019 [1], there are around 58,000 large dams (higher than 15 m) over the world and 75% of them can be classified as embankment dams. Concerning all the constructed dams, the number is much more important. For example, over 91,460 dams were operated across the United States in 2019 [2] and the majority is rock-filled or earth-filled ones. Therefore, safety assessment of embankment dams is crucial for engineers considering their great population and the considerable damages that can be induced by their failures. However, embankment dams involve a high degree of uncertainties, especially for their material properties [3] since they are constructed by natural materials (soils, sands, or rocks), which makes their safety evaluation a difficult task. Probabilistic analysis [4] is an effective solution which permits to rationally account for the soil variabilities and quantify their effects on the dam safety condition by using a reliability

method or a sensitivity method. Additionally, complementary results [5] can be provided by a probabilistic analysis compared to a traditional deterministic assessment, including the failure probability (P_f), design point, model response statistics (e.g., mean and variance) and sensitivity index. Having more results are beneficial for designers to better understand the functioning mode of the dam and make more rational decisions. Therefore, it is worthy to implement probabilistic analyses for the safety assessment of embankment dams, in order to account for the soil variabilities and provide complementary information. **Figure 1** shows a comparison between a probabilistic and a deterministic analysis. In this figure, MCS and FORM [4] are two reliability methods respectively referring to Monte Carlo Simulation and First Order Reliability Method. FoS represents the factor of safety and can be replaced by other types of model responses (like the settlement) which are also of interest for engineers.

In a probabilistic analysis, uncertainties of soil properties can be represented by random variables (RVs) or random fields (RFs) [4]. The former is simpler and easier to couple with a deterministic model [4]. In the RV approach, the soil is assumed to be homogeneous but different values are generated in different simulations for one soil property according to a given distribution. Therefore, the RV method cannot explicitly account for the soil spatial variabilities. On the contrary, the RF approach can model the spatial variation of soils. For a soil property in one simulation, one RF, meaning a collection of different values in a discretized grid, is generated according to the soil parameter statistics and a given autocorrelation structure. However, this approach is more complex and needs extra computational efforts (e.g., quantification of the autocorrelation distances and generation of RFs) compared to the RV one. **Figure 2** illustrates the principle idea of the two approaches.

In this chapter, a practical framework is proposed to efficiently perform the probabilistic analysis of embankment dams. The RV and RF approaches are both implemented into the framework, corresponding to two assessment stages. The RV approach permits a quick estimate on the target results (e.g., P_f) while the RF one is able to account for the soil spatial variability and update the P_f in order to be more precise in a second stage. The proposed framework is applied to an embankment dam stability problem to show its capacity of providing many useful results and its high computational efficiency. A discussion section is provided as well in which the obtained results are validated by comparing with direct MCS. Besides, some issues such as the reliability method selection and probabilistic analysis tools are discussed.

Figure 1.
Contributions of a probabilistic analysis (red color) to a deterministic analysis (blue color).

Figure 2.
 Comparison between the two approaches of uncertainty modeling (left: Random variables; right: Random fields).

2. Presentation of the used probabilistic analysis methods

This section aims to briefly present the probabilistic analysis methods used in the proposed framework including two reliability methods (MCS and FORM), a surrogate modeling technique (PCE), a global sensitivity analysis method (Sobol) and a RF generation approach (KLE).

2.1 Monte Carlo simulations (MCS)

The MCS offers a robust and simple way to estimate the distribution of a random model response and assess the associated Pf . The idea is to largely and randomly generate samples according to a joint input Probability Density Function (PDF) and evaluate the model response of each sample (i.e., an input vector $\mathbf{x}$) by a deterministic computational model. For an MCS with N_{MC} model evaluations, the Pf can be approximated by [4]:

$$Pf \approx \frac{1}{N_{MC}} \cdot \sum_{i=1}^{N_{MC}} I_{MC}(\mathbf{x}^{(i)}) \quad (1)$$

where $I_{MC}(\cdot)$ is an indicator function with $I_{MC}(x) = 1$ if x leads to a failure, otherwise $I_{MC}(x) = 0$. The value of N_{MC} should be large enough in order to obtain an accurate estimate for the Pf which can be assessed by its Coefficient of Variation (CoV):

$$CoV_{Pf} = \sqrt{(1 - Pf)/(N_{MC} \cdot Pf)} \cdot 100\% \quad (2)$$

It is important to mention that the CoV_{Pf} of Eq. (2) is independent of the problem dimension. Additionally, the MCS works regardless of the complexity of the Limit State Surface (LSS). However, a crude MCS suffers from a low computational efficiency. According to Eq. (2), around $100/Pf$ model evaluations are required if the target CoV_{Pf} is 10%.

2.2 First order reliability method (FORM)

The FORM estimates the Pf by approximating the LSS locally at a reference point with a linear expansion. The reference point is called as design point $\mathbf{P}^*$. It is defined in the standard normal space as the point that is on the LSS and closest to

the space origin O_{SN} . This point can be located by solving an optimization problem as [6]:

$$\mathbf{P}^* = \operatorname{argmin}\{\|\mathbf{u}\|, G(\mathbf{u}) \leq 0\} \quad (3)$$

where $\mathbf{u}$ is the input vector $\mathbf{x}$ transformed into the standard normal space and $G(\cdot)$ is the performance function with $G(\mathbf{u}) \leq 0$ representing the failure domain. For the slope stability analysis, the performance function can be defined as: $G = FoS - 1$. Once the $\mathbf{P}^*$ is determined, the Pf can be approximated by the following equation:

$$Pf \approx \Phi_{SN}(-\beta_{HL}) = \Phi_{SN}(-\|\mathbf{P}^*\|) \quad (4)$$

where Φ_{SN} is the standard normal Cumulative Density Function (CDF) and β_{HL} is the Hasofer-Lind reliability index. Additionally, based on the components of the vector from O_{SN} to $\mathbf{P}^*$, the importance factor of each RV can be derived [6].

2.3 Polynomial Chaos expansions (PCE)

The PCE is a powerful and efficient tool for metamodeling which consists in building a surrogate of a complex computational model. It approximates a model response Y by finding a suitable basis of multivariate orthonormal polynomials with respect to the joint input PDF in the Hilbert space. The basic formula of PCE is [7]:

$$Y \approx \sum_{\alpha \in \mathbb{N}^M} k_{\alpha} \Psi_{\alpha}(\xi) \quad (5)$$

where ξ are independent RVs, k_{α} are unknown coefficients to be computed with α being a multidimensional index and Ψ_{α} are multivariate polynomials which are the tensor product of univariate orthonormal polynomials. The representation of Eq. (5) should be truncated to a finite number of terms for practical applications by using the standard or hyperbolic truncation scheme. Then, the unknown coefficients can be estimated by using the Least Angle Regression method. The accuracy of the truncated PCE can be assessed by computing the coefficient of determination R^2 and the Q^2 indicator: R^2 is related to the empirical error using the model responses already existing in the design of experiment (DoE), while Q^2 is obtained by the leave-one-out cross-validation technique [7].

In order to further reduce the number of Ψ_{α} after the truncate operation when the input dimension is high, the sparse PCE (SPCE) was proposed [7]. The idea came from the fact that the non-zero coefficients in the PCE form a sparse subset of the truncation set obtained by the hyperbolic truncation scheme. Thus, it consists in building a suitable sparse basis instead of computing useless terms in the expansions that are eventually negligible.

2.4 Sobol-based global sensitivity analysis (GSA)

The GSA aims to evaluate the sensitivity of a Quantity of Interest (QoI) with respect to each RV over its entire varying range. Among many methods for performing a GSA, the Sobol index has received much attention since they can give accurate results for most models [8]. The Sobol-based GSA is based on the variance decomposition of the model output. The first order Sobol index is given as:

$$S_i = \frac{V_i(Y)}{V_t} = \frac{\operatorname{Var}[\mathbb{E}(Y|x_i)]}{V_t} \quad (6)$$

where V_t is the total variance of Y . For the $Var[\mathbb{E}(Y|x_i)]$, the inner expectation operator $\mathbb{E}(\cdot)$ is the mean of Y considering all possible $x_{\sim i}$ values while keeping x_i constant; the outer variance $Var(\cdot)$ is taken over all possible values of x_i . The first order Sobol index measures the contribution of the variable x_i solely. Another important parameter in a Sobol-based GSA is the total effect index which is given as:

$$S_{Ti} = S_i + \sum_{i \neq j} S_{ij} + \dots S_{1, \dots, M} \quad (7)$$

where $S_{ij}, \dots S_{1, \dots, M}$ represents the higher order Sobol index. S_{Ti} is able to take into account the interaction effects of the variable x_i with other variables.

It is noted that the Sobol index is only effective for independent variables. In order to properly account for the input correlation effect, the Kucherenko index [5] can be employed which is also based on the variance decomposition. For the estimation of the Sobol or Kucherenko index (First order and total effect), the traditional way is to use the idea of MCS however it requires a high number of model evaluations.

2.5 Karhunen-Loève expansions (KLE)

A random field (RF) can describe the spatial correlation of a material property in different locations and represent nonhomogeneous characteristics. The KLE, as a series expansions method, is widely used in the geotechnical reliability analysis since it can lead to the minimal number of RVs involved in a RF discretization [7]. In the KLE context, a stationary Gaussian RF H can be expressed as follows:

$$H(\mathbf{x}_{RF}) \approx \mu + \sigma \sum_{i=1}^{N_{KL}} \sqrt{\lambda_i} \phi_i(\mathbf{x}_{RF}) \xi_i \quad (8)$$

where $\mathbf{x}_{RF}$ is the coordinate of an arbitrary point in the field, μ and σ represents respectively the mean and standard deviation of the RF, λ_i and ϕ_i are respectively the eigenvalues and eigenfunctions of the autocovariance function for the RF, ξ_i is a set of uncorrelated standard normal RVs and N_{KL} is used to truncate the KLE for practical applications. The autocovariance function is the autocorrelation function multiplied by the RF variance. The 2D exponential autocorrelation function is commonly used in the field of reliability analysis. It can be given by:

$$\rho(\mathbf{x}, \mathbf{x}') = \exp \left(-\frac{|\mathbf{x} - \mathbf{x}'|}{L_x} - \frac{|y - y'|}{L_z} \right) \quad (9)$$

where (x, y) and (x', y') are the coordinates of two arbitrary points in the RF, L_x and L_z is respectively the horizontal and vertical autocorrelation distance. The autocorrelation distance is defined as the length which can lead to a decrease from 1 to $1/e$ for the autocorrelation function. Concerning the N_{KL} , its value is determined by evaluating the error due to the truncation term. The variance-based error globally estimated in the RF domain Ω can be expressed as [9]:

$$\varepsilon_{KL} = \frac{1}{\Omega} \int_{\Omega} \left[1 - \sum_{i=1}^{N_{KL}} \lambda_i \phi_i^2(\mathbf{x}_{RF}) \right] d\Omega \quad (10)$$

3. The introduced framework

This section presents the introduced framework for the probabilistic analysis of embankment dams. A flowchart of the framework is given in **Figure 3**.

At the beginning, three elements should be prepared. Firstly, the distribution type and the related parameters (e.g., mean and variance) of the concerned material properties have to be determined. It will allow describing their uncertainties by means of RVs. The selected material properties should be relevant to the QoI of the problem. In case of it is difficult to properly select the relevant properties, all the possible properties can be considered for the RV modeling. The Global Sensitivity Analysis (GSA) which will be performed in the first stage can help to understand the significance of each property. With the GSA results, one can then select which properties will be modeled by RFs. The second work is to develop a deterministic computational model by using numerical or analytical methods (e.g., Finite element method and Limit analysis method). The objective of this model is to estimate the QoI with a given set of input parameters. Then, the autocorrelation structure of the concerned properties should be determined. This structure, defined by an autocorrelation function and the autocorrelation distances, allows to describe the spatial correlation between different locations of a property. It is a key element in the generation of RFs. After these three preparation-works, the analyses in the two stages can be performed. It should be noted that the focus of this chapter is to show the benefits of a probabilistic analysis and demonstrate the proposed framework. Concerning the way of rationally determining the distribution parameters and the autocorrelation structure by using the available measurements, readers can refer to [10, 11].

The objective of the first stage is to provide a variety of probabilistic results with an acceptable computational burden. The results could be helpful to analyze the current problem in a preliminary design phase and guide the following site investigation program and the next design assessment phase. In this stage, the RV approach is used to consider the input uncertainties. It allows quickly having a first

Figure 3.
Flow chart of the proposed framework.

view on the target results given that this approach can be easily coupled with any deterministic model and any probabilistic analysis method. Three analyses are performed in this stage by using respectively three techniques: two reliability methods (MCS and FORM) and one sensitivity method (Sobol-based GSA). The MCS is always considered as a reference method to evaluate other reliability methods due to its robustness. Therefore, an MCS is conducted here in order to obtain an accurate estimate on the P_f . It can also provide the model response distribution and statistics. The FORM is an approximation method due to its linear assumption. It is also adopted in this stage because this method can provide a variety of valuable results which could be beneficial for engineers. For example, the design point permits to know how much margin there are with respect to the current mean values, and the partial safety factors are comparable with the conservative factors used in a deterministic analysis to penalize the strength properties. The Sobol-based GSA permits to quantify the contribution of each RV to the model response variance. By using the Kucherenko index, the correlation effect among the RVs can also be accounted for. According to the GSA results, the properties, which have very slight effects, could be kept as RVs or treated as deterministic in the second stage. This can significantly reduce the computational burden. Particularly, the three analyses are conducted by using a surrogate model (SPCE). The aim is to reduce the total computational time given that a direct MCS or GSA is very time-consuming since they need usually tens of thousands of deterministic model evaluations. For most cases, it is not affordable to repeatedly run a deterministic model with a number higher than 10^4 . Besides, an active learning process [12] given in **Figure 4**, is used to construct the SPCE model. This process starts with an initial DoE and gradually enriches it by adding new samples. A new SPCE model is created each time after the DoE updated with new samples. This process is stopped when some criteria are satisfied. This algorithm is more efficient than the metamodel training based on a single DoE and can give accurate estimate on the P_f .

Figure 4.
An active learning process for the PCE model construction.

Figure 3	1. The procedure of Figure 4 is followed to create an SPCE model
	2. For independent RVs, the Sobol index is used For correlated RVs, the Kucherenko index is used
	3. The RFs are generated by the KLE Conditional RFs can be used if knowing the measurements' locations
	4. The SIR is used to reduce the input dimension An SPCE is created in the reduced space
Figure 4	a. The size N_{ini} of the initial DoE is: $\max [12, M]$ M : input dimension Latin Hypercube Sampling (LHS) is used for generating samples
	b. There is no standard value to check if the problem is high dimensional or not. For the SPCE model, it can easily handle 10–20 RVs. It is thus better to reduce the dimension if $M > 20$ or 30.
	c. An important parameter in the SIR is the slice number N_{sir} $10 \leq N_{sir} \leq 20$ for the cases with several hundred RVs [9] $20 \leq N_{sir} \leq 30$ when the number of input RVs is several thousands
	d. The algorithm presented in [9] is used to create an SPCE The SPCE optimal order is determined by testing in a range
	e. Stopping condition 1 measures if the accuracy indicator Q^2 of the constructed SPCE model is higher than a target value Q_t^2 .
	f. Stopping condition 2 evaluates the convergence of the Pf estimation by computing an error Err_{con} which is the maximum value of the relative errors calculated from all the possible pairs in a vector. The vector consists of the N_{s2} last Pf estimates in the adaptive DoE process. The condition will be satisfied if Err_{con} is lower than a given value Err_t .
	g. N_{add} samples are selected by using the strategy of [12].
	h. An MCS population is generated using the LHS as a candidate pool
	i. DoE is updated by adding the new samples and their model responses

Table 1.
Details remarks to some specific features of Figures 3 and 4.

The second stage aims to consider the spatial variation of the concerned properties which are ignored in the previous stage. It can thus provide a more precise Pf estimate in a second (final) design phase. The new data collected in the new site investigation can be incorporated in this stage in order to update the uncertainty modeling. The GSA results of the first stage can be used to reduce the number of the properties that should be modeled by RFs. The probabilistic analysis becomes a high dimensional problem in this stage due to the RF discretization. As a result, only the MCS is used since the other two methods have difficulties of handling a large number of input RVs (high dimension). The SPCE coupled with the adaptive DoE process is also used at this stage in order to accelerate the MCS. Particularly, a dimension reduction technique – Sliced Inverse Regression (SIR) [9] is used to reduce the input dimension. The SIR is based on the principle that a few linear combinations of original input variables could capture the essential information of model responses. **Table 1** gives a summary of the specific remarks to **Figures 3** and **4**.

4. Application to an embankment dam example

This section shows an application of the proposed framework to an embankment dam stability problem. The dam initially proposed and studied in [5, 13] is selected for this application.

4.1 Presentation of the studied dam and deterministic model

The studied dam is given in **Figure 5**. It has a width of 10 m for the crest and a horizontal filter drain installed at the toe of the downstream slope. The soil is assumed to follow a linear elastic perfectly plastic behavior characterized by the Mohr Coulomb shear failure criterion. In this work, the dam stability issue will be analyzed by considering a constant water level of 11.88 m and a saturated flow. Additionally, a horizontal pseudo-static acceleration of 2.16 m/s^2 toward the downstream part is applied on the dam body. This value represents a relatively high seismic loading and is determined by referring to the recommendations given in [14] for a dam of category A with a soil of type B.

Concerning the input uncertainty modeling, three soil properties (density γ , effective cohesion c' and friction angle ϕ') of the compacted fill are modeled by lognormal RVs or RFs. The illustrative values for the distribution parameters and autocorrelation distances are given in **Table 2**. The uncertainties in the soil hydraulic parameters are not considered since the variation of the dam phreatic level in the downstream part is not significant due to the presence of the filter drain. The mean values are taken from a real dam case reported in [10] and the selected CoVs are consistent with the recommendations give in [3]. A correlation coefficient of -0.3 is considered between c' and ϕ' since it usually exists a negative correlation between these two properties and the correlation coefficient is varied with a range of $[-0.2, -0.7]$ [5]. The L_x is assumed to be significantly larger than L_z since embankment dams are constructed by layers and the spatial variation of material properties is less remarkable in the horizontal direction than the vertical one. The other soil properties are considered as deterministic by using the values given in [5].

The deterministic model used in this work for estimating the dam FoS is developed by using the idea of [13]. It combines three techniques: Morgenstern Price Method (MPM), Genetic Algorithm (GA) and a non-circular slip surface generation

Figure 5.
Geometry of the studied dam (g : gravity acceleration).

Soil property	Distribution	Mean	CoV	Correlation coefficient	L_x (m)	L_z (m)
γ (kN/m ³)	Lognormal	19.8	5%		40	8
c' (kPa)	Lognormal	8.9	30%	-0.3	40	8
ϕ' (degree)	Lognormal	34.8	10%		40	8

Table 2.
Distribution parameters and autocorrelation distance for γ , c' and ϕ' .

method. MPM is employed to compute the FoS of a given failure surface; GA aims at locating the most critical slip surface (i.e., minimum FoS) by performing an optimization work; The implementation of non-circular slip surfaces can lead to more rational failure mechanics for the cases of non-homogeneous soils. The principle of the model is to firstly generate a number of trial slip surfaces as an initial population, and then to determine the minimum FoS value by modeling a natural process along generations including reproduction, crossover, mutation and survivors' selection. The distribution of the pore water pressures inside the dam is given by a numerical model [5]. In this work, the developed deterministic model is termed as LEM-GA. Using a simplified deterministic model (e.g., LEM-GA) is beneficial for a reliability analysis since it can reduce the total computational time. This strategy can thus be adopted in a preliminary design/assessment phase for efficiently obtaining first results. Then, a sophisticated model (e.g., Finite element model) is required in a next phase if complex conditions should be modeled (e.g., rapid drawdown and unsaturated flows) or multiple model responses (e.g., settlement and flow rate) are necessary.

4.2 First stage: RV approach

This section shows the conducted works at the first stage of the proposed framework and presents the obtained results. The RV approach is used in order to have a quick estimate on the dam reliability and the contribution of each input variable. The joint input PDF is defined by the mean, CoV and $\beta_{c\phi}$ of the three soil properties given in **Table 2**. Three probabilistic analyses (MCS, FORM, and GSA) are performed with a surrogate model, also known as meta-model, in this stage so that a variety of useful results can be obtained.

Firstly, an SPCE surrogate model is constructed as an approximation to the model LEM-GA. It is achieved by using the procedure of **Figure 4** with the following user-defined parameters: $Q_t^2 = 0.98$; $Err_t = 0.1$; $N_{s2}=10$; $N_{add} = 1$; $N_{ini} = 12$; the size for the MCS candidate pool is large enough so that the estimated Pf has a CoV lower than 5%. The finally obtained SPCE is a 3-order model with a Q^2 of 0.99. Twelve new samples, determined by the active learning process, are added to the initial DoE which corresponds to a total number of model evaluation (N_{me}) of 24. Then, the SPCE model is respectively coupled with MCS, FORM and GSA in order to provide different results. As a meta-model is usually expressed analytically, the SPCE-based analyses are thus very fast. Therefore, the main computation burden of the first stage lies in constructing a satisfactory SPCE model. In this work, only 24 deterministic calculations are performed for the construction, representing a significant reduction in N_{me} compared to direct MCS, FORM and GSA which require at least tens thousands of model evaluations. This shows the main advantage of the first stage in the proposed framework: benefiting from the computational efficiency of a meta-model and providing a variety of useful results.

Figure 6 presents the results provided by the SPCE-aided MCS with $N_{MC} = 10^5$. According to the obtained 10^5 FoS values, its PDF and CDF can be plotted. The PDF shows that the dam possible FoS under the current calculation configuration mainly varies between 1 and 1.6 with a mean (μ_{FoS}) of 1.285 and a standard deviation (σ_{FoS}) of 0.137. Giving the CDF allows approximately estimating the probability of getting a FoS lower than any threshold. Then, the dam Pf is obtained by computing the ratio between two numbers: N_f and N_{MC} with N_f representing the number of the FoS values lower than 1 (i.e., failure). **Figure 7** presents the results obtained by the SPCE-aided FORM and GSA. The FORM is an approximation method by the fact that it assumes a linear expansion tangent to the LSS at P^* for the Pf estimation.

Figure 6.
PDF and CDF of the FoS values obtained in the first stage.

The advantage of the FORM is that it is able to give many results in terms of reliability index (β_{HL}), design point, partial safety factor (FoS) and importance factor of each variable. The design point represents the most probable failure point in the FORM context, and can be used together with the partial FoS to guide a deterministic analysis on the same problem. The GSA aims to quantify the contribution of each soil property, modeled by RV, with respect to the dam FoS variance. The results permit to make a rank of all the variables according to their importance as shown in **Figure 7**. The Kucherenko index is used here for the GSA since there exists a correlation between the input variables. The total effect index considers both the independent impact of one variable and its correlation effect with other variables. According to **Figure 7**, it is observed that the variable ϕ' is dominant for the FoS variation under the current probabilistic input configuration (**Table 2**).

Figure 7.
GSA and FORM results of the dam in the first stage.

The variable c' has also a noticeable contribution while the γ effect is very slight. It should be noted that the importance factor (FORM) and the sensitivity index (Sobol or Kucherenko -based GSA) are different between each other. The former measures the contribution of a RV with respect to the failure while the latter quantifies the importance of a RV regarding the QoI variation. Additionally, the importance factor by FORM holds only for the case with independent RVs. The related results are still given in **Figure 7** in order to have a rank and to compare with the GSA estimates.

In summary, this stage provides a first estimate on the dam P_f which can be used to evaluate the design of a new dam or the safety condition of an existing dam. The other information, such as the FoS statistics and design points, are also helpful for this first evaluation. The sensitivity analysis results permit to know the contribution of the considered soil properties and treat their uncertainties with different ways in a next verification/design phase.

4.3 Second stage: RF approach

The second stage of the proposed framework is to consider the soil spatial variability by RFs and obtain a more precise P_f estimate. According to the results of the first stage, the effect of the variable γ is almost negligible for the dam failure or the FoS variance. Therefore, it is reasonable to only model c' and ϕ' by RFs and keep representing γ by RVs in the second stage. This can make the analysis of this stage simpler and faster given that generating RFs and mapping them to a model require extra computational efforts. Besides, the input dimension can be reduced compared to considering three RFs (c' , ϕ' , and γ) for each simulation since there is no need to do the γ discretization. In this stage, the c' and ϕ' are modeled by cross-correlated lognormal RFs using the parameters of **Table 2**, while the γ is treated as same as the previous stage.

The first step in this stage is to determine the truncation term number N_{KL} once the necessary probabilistic parameters (mean, CoV, L_x and L_z) are defined. It can be achieved by evaluating the truncation error of a KLE RF evaluated by Eq. (10) with a target accuracy. In this work, the N_{KL} is determined for a ε_{KL} lower than 5%.

Figure 8.
RF truncation error and example of the generated RFs.

Figure 8 plots the ε_{KL} against the N_{KL} and finally a $N_{KL} = 125$ is adopted for the case of $L_x = 40m$ and $L_z = 8m$ (**Table 2**). Then, the input dimension for the reliability analysis in this stage is 251 since two RFs (c' and ϕ') and one RV (γ) should be considered for each simulation. In **Figure 8**, an example of the c' RF generated by the KLE with the pre-defined parameters is illustrated. It can be seen that c' varies more significantly in the vertical direction than the horizontal one.

The second step is to create an SPCE model to replace the LEM-GA coupled with RFs. The active learning process of **Figure 4** is followed for the SPCE training with the user-defined parameters given as: $Q_t^2 = 0.98$; $Err_t = 0.15$; $N_{s2}=5$; $N_{add} = 2$; $N_{ini} = 251$; the size for the MCS candidate pool is large enough so that the estimated Pf has a CoV lower than 5%. Additionally, the input dimension is reduced by using the SIR *a priori* the SPCE construction at each iteration with the current DoE. This is because that the considered reliability analysis is a high dimensional problem which has 251 input RVs. Directly training an SPCE with the original input space will require a large size of DoE and may lead to a less accurate meta-model. By performing an SIR with a slice number of 20, the input dimension is reduced from 251 to 19. Then, it is possible to create an SPCE model with respect to the 19 new RVs using an acceptable size of DoE (e.g., several hundred). At the end, the obtained SPCE is a 2-order model with a Q^2 of 0.99. The final size of the DoE is 423 which means that 172 new samples are added in the adaptive process in order to improve the SPCE performance in estimating the dam Pf .

The last step is to perform an MCS with the determined SPCE model. The obtained results are presented in **Figure 9**. The dam FoS mainly varies between 1.1 and 1.5 with a mean of 1.276 and a standard deviation of 0.102. The dam Pf is estimated as 6×10^{-4} . Compared to the analysis of the first stage, the current analysis leads to a clearly reduced σ_{FoS} corresponding to a narrower variation range as shown by the PDF. The dam Pf is also decreased by around one order of magnitude. The comparison between **Figures 6** and **9** indicate that using RFs instead of RVs to model the soil variabilities can reduce the FoS uncertainty and provide a lower Pf estimate. Although considering the soil spatial variability requires extra computational efforts for RFs generation and makes the reliability

Figure 9.
MCS results of the dam obtained at the second stage.

analysis more complex, it is worthy to do so since a more precise Pf estimate can be obtained and can lead to a more economic design. A detailed explanation about the Pf decrease from the RV to RF approach will be given later.

4.4 Parametric analysis

It needs in some cases to perform a series of parametric analyses. The objective is to evaluate the effects of some parameters which are difficult to be precisely quantified due to the lack of enough measurements. The physical range recommended in literature for the concerned parameters can be used to define some testing values. In the proposed framework, the computational burden for conducting such parametric analyses is acceptable since the use of the SPCE model significantly reduces the consuming time of one probabilistic analysis. In this work, the effects of two parameters on the dam reliability are investigated: the cross-correlation between c' and ϕ' ($\beta_{c\phi}$) and the vertical autocorrelation distance (L_z). In the reference case (**Table 2**), the $\beta_{c\phi}$ is assumed to be -0.3 . In this section, two testing values (0 and -0.6) are selected for the $\beta_{c\phi}$ to check its influence: $\beta_{c\phi} = 0$ represents independent input RVs while $\beta_{c\phi} = -0.6$ is a strongly correlated case. Then for the L_z , two values (40 and 3 m) are adopted as two complementary cases to the assumed L_z in the reference case (8 m). $L_z = 40$ m leads to isotropic RFs given that L_x is also 40 m and represents a relatively homogenous soil, while $L_z = 3$ m allows to consider a soil significantly varying along depth. The L_x is assumed to be constant with 40 m in this case. Such an assumption is based on the fact that embankment dams are usually constructed by layers leading to highly correlated soil properties in the horizontal direction if the construction materials are well selected. **Table 3** gives a summary of all the cases considered in this section. Case 1B and 2B in this table refers to the reference case which is performed respectively in the first and second stage of the previous sections. In the RV approach, the soil is assumed to be homogenous which means that the values of different locations in this field are perfectly correlated. Therefore, this approach corresponds to an infinite L_x and L_z . The input dimension in **Table 3** means the number of input RVs for each case. The dimension is 3 for all the cases with the RV approach which represents the three soil properties (γ , c' and ϕ'). For the RF approach, the dimension is related to the truncation term N_{KL} as determined in **Figure 8**. The N_{KL} should be increased if smaller L_x or L_z are considered. In other words, it means than an accurate representation of a RF with small autocorrelation distances requires more RVs.

Figure 10 presents the obtained results of the parametric analysis (1A, 1B, and 1C) for the $\beta_{c\phi}$ effect. The SPCE is used for the meta-model construction and it is coupled only with MCS since the focus here is to estimate the dam Pf . From this figure, it is observed that the FoS PDF becomes taller and narrower when the $\beta_{c\phi}$ is

	Case	Approach	Distribution parameters	$\beta_{c\phi}$	L_x (m)	L_z (m)	Input dimension
$\beta_{c\phi}$ Effect	1A	RV	Table 2	0	∞	∞	3
	1B	RV	Table 2	-0.3	∞	∞	3
	1C	RV	Table 2	-0.6	∞	∞	3
L_z Effect	2A	RF	Table 2	-0.3	40	40	61
	2B	RF	Table 2	-0.3	40	8	251
	2C	RF	Table 2	-0.3	40	3	621

Table 3.
The analyses conducted in this section.

decreased from 0 to -0.6 . The PDF of the independent case leads to the most scattered FoS values. Consequently, the dam P_f estimate, being the tail probability of a distribution, is decreased from Case 1A to 1C. The P_f decrease corresponds to a change of one order of magnitude when $\beta_{c\phi}$ is reduced from 0 to -0.6 . Considering a negative cross-correlation between c' and ϕ' can reduce the total input uncertainty. Therefore, the output variance can also be reduced given that these two properties are dominant for the FoS variation according **Figure 7**. Additionally, the number of small FoS values is decreased since a negative $\beta_{c\phi}$ can partially avoid generating a small value for both c' and ϕ' in one simulation, which then leads to a lower P_f .

Figure 11 shows the results for the investigation on the L_z effect. The results of Case 1B are presented as well in this figure which permits a comparison between the RV and RF approach. The SPCE-aided MCS is used by following the algorithm of **Figure 4** to perform the reliability analysis. Particularly, the input dimension is reduced by using the SIR each time before the SPCE construction because the three considered cases (2A, 2B, and 2C) are high dimensional problems due to the RF discretization as shown in **Table 3**. It can be observed that the FoS PDF is taller and narrower with decreasing the L_z . This means that a smaller L_z can lead to a FoS with less uncertainty. As a result, the tail probability of the distribution (P_f) is also decreased from Case 2A to 2C with a change of two orders of magnitude showing that the L_z effect is remarkable on the dam P_f . The RV approach provides the most scattered FoS distribution and the highest P_f . A possible explanation for these findings is given as below. Two RFs for respectively Case 2A and 2C are generated and presented in **Figure 12** in order to help the following interpretation. A large L_x or L_z value means a great probability of forming large uniform areas as shown in **Figure 12** (upper part). The global average of the field could be low, medium or high which means a large variation for the global average among different realizations of RFs. The global average is partially related to the estimated FoS so the latter could also have a large variation as evidenced in **Figure 11**. Then, the P_f is higher since it is the tail probability of a distribution. On the contrary, for the case with a

Figure 10.
Comparison of the PDF and P_f by considering different $\beta_{c\phi}$.

Figure 11.
Comparison of the PDF and Pf considering different L_z .

Figure 12.
Example of the RFs generated in Case 2A and 2C.

low L_x or L_z value, there are probably some relatively higher values generated close to the area with low values and vice versa. As a result, the global average varies in a narrower range also the FoS, so the Pf is lower. Additionally, the failure surface seeks the weak areas, so it is in general longer and less smooth when L_x and L_z are small. For a long and rough slip surface, more energies are required for its movement which means a relatively high FoS. Therefore, the Pf is lower with small L_x and L_z . As these two parameters are assumed to be infinite in the RV approach, the largest uncertainty in the FoS and the highest Pf are obtained.

5. Discussions

5.1 Validation of the surrogate-based results

The proposed framework is based on the metamodeling to perform a probabilistic analysis. Therefore, the key element of the proposed framework is to create an accurate SPCE model which can well replace the original computational model. In the next paragraph, two recommendations are given for a good metamodeling.

Firstly, it is recommended to use a space-filling sampling technique (e.g., LHS) to generate samples from a given PDF for the initial DoE and the MCS candidate pool. This allows generating a set of samples which can reasonably cover the input space. The LHS is also faster than a purely random sampling technique for the result convergence in an MCS. Secondly, an active learning process, such as the one of **Figure 4**, is highly suggested for the SPCE construction. The process is stopped only if stable P_f estimates are reached and the added samples in this process are those which can improve the SPCE performance in predicting failures. Therefore, one can have more confidence on the obtained P_f estimate by using this process. Besides, the DoE is gradually enriched until the stopping conditions are met. As a result, the size of the DoE can be automatically determined, and the issue of overfitting may be avoided.

Concerning the validation of the constructed surrogate model, three solutions are provided here. The first one is to use the available results in the DoE to compute an accuracy indicator for the meta-model, such as the Q^2 in the PCE. The Q^2 is obtained by the leave-one-out error which is a type of the k -fold cross validation techniques. The advantage of this solution is that no complementary model evaluations are required, and the current DoE is fully explored. Then, the second solution is to use a validation set in which new samples, not covered in the current DoE, are generated and evaluated by both the surrogate and deterministic models. The predictions made by the two models for the new samples can be compared in order to check the accuracy of the obtained meta-model. The new samples can be obtained randomly by the LHS or selected close to the LSS so that the meta-model capacity in classifying safe/failure samples is then verified. The third solution involves performing a direct MCS, FORM or GSA to validate the results obtained by the surrogate-aided analyses. Obviously, this solution requires a huge computational effort if a direct MCS or GSA should be conducted which means that no surrogate model is used, and MCS/GSA is directly coupled with the original computational model. Therefore, it is not an applicable solution for all cases. It could be effective when a series of analyses are performed so a direct MCS can be used to validate one analysis.

In this section, the third validation solution is adopted since some parametric analyses are carried out and the employed deterministic model (LEM-GA) is not too time-consuming. Two cases (1B and 2A) are selected for the validation and are analyzed by a direct MCS in this section. The N_{MC} in the direct MCS is determined so that the CoV_{P_f} is around 10%. **Figure 13** compares the FoS PDF of the two analyses obtained by the two methods (SPCE-aided MCS and direct MCS). It clearly shows that the two PDF curves of the two methods are almost superposed with each other for both the two cases. This indicates a good approximation of the SPCE to the original model.

Table 4 gives a detailed comparison between the two methods in terms of P_f , FoS statistics and computational efficiency. It is found that the P_f of SPCE-MCS is close to the reference result (direct MCS) with an error lower than 6% for both the two cases. The 95% confidence bounds of the P_f estimates are also given in this

Figure 13.
Comparison of the PDF provided by the two methods.

Case	Method	$P_f (\times 10^{-3})$			FoS		Computational efficiency	
		Mean	Low ⁽¹⁾	Up ⁽¹⁾	μ_{FoS}	σ_{FoS}	N_{me}	T_{tc}
1B	Direct MCS	7.50	6.07	8.93	1.29	0.14	14,000	25 hours
	SPCE-MCS	7.94	7.39	8.49	1.28	0.14	24	3 minutes
2A	Direct MCS	2.80	2.25	3.35	1.28	0.12	35,000	63 hours
	SPCE-MCS	2.84	2.69	2.99	1.28	0.12	131	25 minutes

⁽¹⁾95% confidence bounds of the P_f estimate.

Table 4.
Comparison of the reliability results and computational efficiency provided by the two methods.

table. If the MCS size is large enough, the estimated P_f can be approximated by a normal distribution which makes the confidence bounds to be available. It is found that the P_f confidence bounds of SPCE-MCS are covered by the ones of the direct MCS. In a surrogate-aided reliability analysis, it is acceptable to largely increase the MCS size in order to obtain a small CoV_{P_f} . However, much more computational efforts are required in the direct MCS if its size should be enlarged, so a CoV_{P_f} of around 10% is adopted in this work. This finding and argument mean that a precise P_f with a small CoV_{P_f} could be easily obtained using the proposed framework. Then, for the FoS statistics, the two methods show a good agreement between each other. This is not surprising since the two methods provide closely similar FoS distribution as evidenced in **Figure 13**. Concerning the computational efficiency, two terms are presented in **Table 4** for a comparison: N_{me} (number of deterministic model evaluations) and T_{tc} (total computational time). The T_{tc} is evaluated in a computer equipped with an CPU of Intel Xeon E5-2609 v4 1.7 GHz (2 processors). It is observed that the N_{me} is significantly reduced by using the SPCE compared to a

direct MCS (e.g., from 14,000 to 24 for Case 1B), corresponding to a considerable reduction in the T_{tc} (from 25 hours to 3 minutes). Due to the computational efficiency of the SPCE-aided MCS, it is then possible to carry out some parametric analyses in order to investigate the effects of some parameters in a probabilistic framework. The necessary size of the DoE to construct a satisfactory SPCE model is dependent of the input dimension. In general, a higher dimension requires more model evaluations for the SPCE training.

5.2 Practical applications

This section provides a discussion on some issues of a probabilistic analysis. The objective is to help engineers to better implement the proposed framework into practical problems.

5.2.1 Probabilistic analysis tool

Probabilistic analysis has received much attention during the last decade in literature. However, it is still not widely applied in practical engineering problems. One major reason which hinders its application in practice is the complexity of performing a probabilistic analysis including understanding/programming a reliability method, RF generation and couple them with a deterministic model. This problem is being addressed in recent years with the establishment of many probabilistic analysis tools. A variety of reliability/sensitivity methods are available in these tools and can be linked with a computational model developed in a third-party software. Examples of these tools include UQlab in Matlab and OpenTURNS in Python. A review of the structural reliability analysis tools can be found in [15]. Using a well-checked tool to perform the probabilistic analysis of practical engineering problems can also avoid personal programming mistakes which could lead to inaccurate results.

5.2.2 Reliability method selection

The proposed framework is based on the SPCE surrogate model. The SPCE is adopted since it has been widely and successfully used in many studies of geotechnical reliability analysis [9, 13, 16]. Some techniques were proposed to be coupled with SPCE in order to efficiently consider the cases with RFs [17], so the SPCE can also handle high dimensional stochastic problems. However, the proposed framework is not limited to the SPCE. It can be updated by using another metamodeling technique (e.g., Kriging and Support Vector Machine) with some necessary modifications. Besides, for estimating a very low P_f (e.g., $<10^{-6}$), the SPCE-aided MCS could be time-consuming given that generating a great number of samples (e.g., $N_{MC} > 10^8$) and operating them requires a big memory in a PC. To tackle this problem, the SPCE can be coupled with other reliability methods in order to alleviate the computational burden. The Subset Simulation (SS) [6, 18] is a good choice to replace the MCS for the above-mentioned case, because SS is independent of the input dimension and LSS complexity.

5.2.3 Parameter selection

This chapter focuses on presenting the proposed framework and showing its application to a dam problem. The soil variability modeling is not explained in detail. How to properly describe the soil uncertainties by using a limited number of

measurements is also an important element for geotechnical probabilistic analysis in practice. Some studies on this topic can be found in [10, 11]. In this chapter, the effects of two parameters ($\beta_{c\phi}$ and L_z) on the dam reliability are discussed by performing two parametric analyses. Both of them have a significant influence on the dam P_f which is decreased with decreasing $\beta_{c\phi}$ or L_z . It seems then logical to use higher values (e.g., $\beta_{c\phi}=0$ and $L_z=40$ m) in order to obtain conservative results if their values cannot be precisely quantified. Attention must be paid for the selection of L_z or L_x because some recent studies [19] demonstrate that it may exist a worst L_z or L_x which can lead to the highest P_f . Therefore, it is advised to perform a parametric analysis on these parameters in order to avoid unsafe designs.

5.2.4 Extension of the proposed framework

The illustrative example in this chapter is based on the stability problem of a homogeneous embankment dam. The proposed framework can also be easily extended to perform the probabilistic analysis of other problems in dams engineering (rapid drawdown, erosion and settlements) by using an appropriate deterministic model and well determining the input uncertainties. Then, the proposed two stages of RV and RF can be conducted in a hierarchical way. For embankment dams with an earth core or multiple soil layers, the uncertainties should be separately modeled for each zone using different RVs or RFs [17]. It is also important to consider the correlation between the variable properties of different zones by analyzing the available measurements. In case of not enough data, a parametric analysis is recommended in order to have an idea of the unknown correlation structure effect. As embankment dams are artificial rock-filled or earth-filled structures constructed with a careful control, uncertainties at the zone boundaries can be considered as negligible. In natural soils, where stratigraphic boundary uncertainties are expected to exist, the related effects will be noticeable and should be considered.

6. Conclusion

This chapter introduces a framework for the probabilistic analysis of embankment dams. The proposed methodology can also be used for other geotechnical works. The RV and RF approaches are both considered in the framework, corresponding two probabilistic analysis stages. In the first stage, the RV approach is used within three probabilistic techniques (MCS, FORM, and GSA) in order to efficiently provide multiple results which could be beneficial for evaluating a design and guide a further site investigation or a further analysis. The second stage introduces RFs for the purpose of accounting for the soil spatial variability and giving a more precise P_f estimate. The metamodeling technique, SPCE, is used in both the two stages aiming to alleviate the total computational burden. Particularly, an active learning process is adopted to construct the required SPCE model. This can further reduce the calculation time of a probabilistic analysis and improve the SPCE accuracy in estimating P_f . The proposed framework is applied to an embankment dam stability problem. A variety of interesting results for the dam considering the soil uncertainties are obtained. The results include the P_f , FoS statistics/distribution, sensitivity index of each soil property, design point and partial safety factors. The provided results (P_f and FoS values) are validated by comparing with a direct MCS. The validation also highlights the efficiency of the introduced reliability method which can reduce the total computational time from several days to less than 1 hour for the two considered cases.

IntechOpen

Author details

Xiangfeng Guo¹ and Daniel Dias^{1,2,3*}

¹ 3SR laboratory, Grenoble Alpes University, France

² Antea Group, Antony, France

³ Hefei University of Technology, China

*Address all correspondence to: daniel.dias@3sr-grenoble.fr

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] ICOLD. World Register of Dams. 2020. Available from: <https://www.icold-cigb.org/GB/icold/icold.asp> [Accessed: 01 April 2020]
- [2] Hariri-Ardebili M-A, Salamon J, Mazza G, Tosun H, Xu B. Advances in dam engineering. *Infrastructures*. 2020;5:39. DOI: 10.3390/infrastructures5050039
- [3] Phoon K-K, Kulhawy FH. Characterization of geotechnical variability. *Canadian Geotechnical Journal*. 1999;36:612-624. DOI: 10.1139/t99-038
- [4] Griffiths D, Fenton G. Probabilistic Methods in Geotechnical Engineering. Vienna: Springer Vienna; 2007. DOI: 10.1007/978-3-211-73366-0
- [5] Guo X, Dias D. Kriging based reliability and sensitivity analysis – Application to the stability of an earth dam. *Computers and Geotechnics*. 2020;120:103411. DOI: 10.1016/j.compgeo.2019.103411
- [6] Bourinet J-M. Reliability Analysis and Optimal Design under Uncertainty Focus on Adaptive Surrogate-Based Approaches. Clermont-Ferrand: Université Clermont Auvergne; 2018
- [7] Sudret B. Polynomial chaos expansions and stochastic finite element methods. In: Phoon K-K, Ching J, editors. Risk and Reliability in Geotechnical Engineering. Florida: CRC Press; 2014. pp. 265-300
- [8] Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*. 2008;93:964-979. DOI: 10.1016/j.res.2007.04.002
- [9] Pan Q, Dias D. Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions. *Reliability Engineering and System Safety*. 2017;167:484–493. DOI: 10.1016/j.res.2017.06.026
- [10] Mouyeaux A, Carvajal C, Bressolette P, Peyras L, Breul P, Bacconnet C. Probabilistic stability analysis of an earth dam by stochastic finite element method based on field data. *Computers and Geotechnics*. 2018;101:34–47. DOI: 10.1016/j.compgeo.2018.04.017
- [11] Papaioannou I, Straub D. Learning soil parameters and updating geotechnical reliability estimates under spatial variability–theory and application to shallow foundations. *Georisk*. 2017;11:116–128. DOI: 10.1080/17499518.2016.1250280
- [12] Marelli S, Sudret B. An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. *Structural Safety*. 2018;75:67–74. DOI: 10.1016/J.STRUSAFE.2018.06.003
- [13] Guo X, Dias D, Pan Q. Probabilistic stability analysis of an embankment dam considering soil spatial variability. *Computers and Geotechnics*. 2019;113:103093. DOI: 10.1016.j.compgeo.2019.103093
- [14] Loudière D, Hoonakker M, Le Delliou P. Risque sismique et sécurité des ouvrages hydrauliques. Paris: Ministère de l'Écologie de France; 2014 (in French)
- [15] Chehade FEH, Younes R. Structural reliability software and calculation tools: A review. *Innovative Infrastructure Solutions*. 2020;5:1–16. DOI: 10.1007/s41062-020-0282-4
- [16] Al-Bittar T, Soubra A-H. Efficient sparse polynomial chaos expansion methodology for the probabilistic analysis of computationally-expensive

deterministic models. International Journal for Numerical and Analytical Methods in Geomechanics. 2014;**38**: 1211–1230. DOI: 10.1002/nag.2251

[17] Guo X, Dias D, Carvajal C, Peyras L, Breul P. A comparative study of different reliability methods for high dimensional stochastic problems related to earth dam stability analyses. Engineering Structures. 2019;**188**:591–602. DOI: 10.1016/j.engstruct.2019.03.056

[18] Li HS, Cao ZJ. Matlab codes of subset simulation for reliability analysis and structural optimization. Structural and Multidisciplinary Optimization. 2016;**54**:391–410. DOI: 10.1007/s00158-016-1414-5

[19] Zhu D, Griffiths DV, Fenton GA. Worst-case spatial correlation length in probabilistic slope stability analysis. Géotechnique. 2018;**69**:85–88. DOI: 10.1680/jgeot.17.t.050