Soils and Rocks
An International Journal of Geotechnical and Geoenvironmental Engineering
www.soilsandrocks.com
ISSN 1980-9743
ISSN-e 2675-5475

Geostatistical-based enhancement of RFEM regarding reproduction of spatial correlation structures and conditional simulations

Jean Lucas dos Passos Belo[1#], Paulo Ivo Braga de Queiroz[1],
Jefferson Lins da Silva[2]

Article

Keywords
Spatial variability
Conditional simulation
Probabilistic analysis
Geostatistics
Random finite element
Reliability

Abstract

Engineering always deals with uncertainties, and efforts are needed to quantify them. A probabilistic analysis considers the statistical information of the problem to this quantification. In the geotechnical area, uncertainties play a particular role in structure design because it deals with naturally formed materials. Evaluating spatial variability has become progressively important. However, studies on the correct reproduction of this variability and conditional simulations are limited. In this paper, a geostatistical-based enhancement of the Random Finite Element Method (RFEM) is presented. The main aim of this study is to incorporate an advanced multivariate geostatistical technique (i.e., Turning Bands Co-simulation, TBCOSIM) to reproduce the coregionalization model of soil properties correctly in order to investigate the effects regarding this reproduction. It is illustrated in a real case of soil slope. The results showed that, for the unconditional simulation, the presented approach reached a perfect agreement with the coregionalization model, while the conditional simulation inserted some disturbances to this agreement, but it still satisfactorily reproduced the model. The original RFEM failed to reproduce this structure, leading to lower variances than the presented approach, which would cause a non-conservative design. Furthermore, disregarding the local uncertainty (i.e., the nugget effect) may introduce bias to analysis and, depending on its magnitude, may also lead the conditional analysis to not show a worthwhile reduction in variances of results. Finally, this paper shows that correctly determining the coregionalization model and reproducing it on probabilistic analysis may meaningfully influence the results.

1. Introduction

Although we (engineers) are typically used to considering engineering as an exact science, we do not always treat it as such. All engineering areas deal with uncertainties (e.g., inherent, spatial, temporal, from measurements or from a model), however we do not always take them into account. Considering and quantifying these uncertainties enables us to evaluate the precision threshold of our estimations (calculated results). When engineers understand the importance of uncertainty quantification and start considering it in their designs and analysis, only then will engineering be conducted as an exact science, within its limitations. Efforts to determine these thresholds should not be overlooked, such as currently observed in practice and even in academic applications, unfortunately. Therefore, this topic requires due attention and may lead to a lengthy discussion.

For Geotechnical Engineering, uncertainties and variabilities associated with material properties, which make up a geotechnical structure, have substantial influences on its safety and behavior. This sensitivity is significant in this area of study because it deals with naturally formed materials (i.e., soils and rocks), sources of large variances and heterogeneity.

Nowadays, many commercial programs allow the realization of probabilistic analysis to evaluate geotechnical structures. Usually, these programs apply the Monte Carlo simulation (MCS) technique, associated with the Limit Equilibrium Method (LEM) or the Finite Element Method (FEM), to perform this analysis. However, today, these programs hold limited resources, such as the number of random variables, the type of probability density functions, the spatial variability consideration, the high computational cost, among others (Belo & Silva, 2020; Belo et al., 2022).

[#]Corresponding author. E-mail address: belo.jeanlucas@gmail.com
[1]Instituto Tecnológico de Aeronáutica, Departamento de Infraestrutura Aeroportuária, São José dos Campos, SP, Brasil.
[2]Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, SP, Brasil.

Submitted on October 23, 2021; Final Acceptance on August 3, 2022; Discussion open until February 28, 2023.
https://doi.org/10.28927/SR.2022.076121

In the literature, studies have proposed various probabilistic approaches for the geotechnical area. The Random Finite Element Method (RFEM), proposed by Griffiths & Fenton (1993), is the most accepted and used approach for this purpose. RFEM reconciles the FEM with the Random Field Theory (RFT) to simulate the spatial variability of soil properties. It correctly searches for the weakest path through heterogenous material and leads to probabilities of failure higher than would be estimated by disregarding the spatial variability (Sayão et al., 2012). Nevertheless, RFEM disregards known field data, usually determined by sampling material or site investigation. It is the major disadvantage of this approach and, disregarding these data and their positions within the field, can cause fluctuations in variance of analysis and hence produce an incompatible design, touching safety and economy. Therefore, its application (unconditional to data) is useful in generic geotechnical design, e.g., in terms of code or standard provisions.

Approaches aiming to consider known field data make use of the Conditional Random Field (CRF). Despite the fact that studies regarding CRF have increased in recent years, studies using this technique in geotechnical engineering remain limited in number. Mrabet & Bouayed (2000) used the CRF to reduce the uncertainty on the probabilistic results of a dam analysis, more specifically regarding the properties of the compacted soil masses. Folle et al. (2006) presented the main statistical and geostatistical methods in geotechnics to deal with quantification regarding the heterogeneity of soil properties. Then, they evaluated a case study using Sequential Gaussian Simulation (SGS). Griffiths et al. (2009) investigated the influence of spatial variability on slope reliability using the RFEM. However, the spatial correlation function was assumed to be fixed, described by an exponentially decaying (Markov) function. Monteiro et al. (2009) approached the problem of rock characterization using drill measurements. The authors incorporated the spatial relationship using the CRF to infer the geology of the neighboring regions. Kim & Sitar (2013) applied the CRF to a homogeneous soil slope to investigate its stability, assuming the deterministic critical slip surface as fixed for the probabilistic analysis.

Lately, Schöbi & Sudret (2015) combined the CRF with the framework of sparse polynomial chaos expansions to analyze response quantities in geotechnical problems, illustrated by applying the approach to a strip foundation problem on a two-layer soil mass. Li et al. (2016) presented a method that combined 3D kriging with a random field generator to develop the CRF. The authors applied this for a slope stability analysis, aiming to identify the best locations for site investigations and compare different candidate slope designs. Liu et al. (2017) applied the CRF to investigate a cohesion-frictional slope using a MATLAB developed code. The CRF was generated using the kriging method and the Cholesky decomposition technique. Yang et al. (2017) used the CRF to investigate undrained slope stability based on the RFEM and a kriging method. Despite not considering conditional simulations, Muñoz et al. (2018), as Griffiths et al. (2009), investigated the influence of spatial variability of the soil parameters on the factor of safety (FoS) of a hypothetical slope. However, the analysis assumed no correlations between variables (univariate analysis, i.e., independent random variables), spatial variability followed normal and lognormal distribution, by using Monte Carlo simulations and Kriging process, and found the FoS with the LEM. Johari & Gholampour (2018) developed a MATLAB code to apply the CRF to a stochastic analysis of an unsaturated soil slope. Yang et al. (2019) used the CRF to investigate the “optimal” site investigation scope for a slope design, combining the analysis of the cost of site investigation with the cost of slope failure. Johari & Fooladi (2020) presented a probabilistic analysis of a real soil slope using the concepts of the CRF, coded in MATLAB. Jurado et al. (2020) proposed a rational approach to test the spatial variance of soil based on site investigation and on the CRF.

Incorporating the data known from the field and its spatial positions in the analysis can be performed by using geostatistical concepts and techniques. Geostatistical simulations enable the generation of random fields that agree with their statistical information and eventual conditioning data. Liu et al. (2019) showed that, among the random field generation methods, seven cover most studies, i.e., sequential Gaussian simulation (SGS), local average subdivision (LAS), turning bands simulation (TBS), spectral method (SM), Karhunen-Loève expansion (KLE), matrix decomposition method (MD) and moving average method (MA).

KLE and MD present the highest algorithm complexity, followed by MA, SM, TBS, SGS and LAS. Furthermore, the algorithm complexity can be powered in cases of multivariate applications. Although KLE has been widely used in stochastic approaches, it poses a problem in applications, where complex geometry will be encountered or when assuming high dimension covariance matrix, and some problems have been identified regarding heterogeneity of the generated sample functions (Sudret & Der Kiureghian, 2000; Stefanou & Papadrakakis, 2007). MD, as it is, suffers from several deficiencies, e.g., like KLE, problems with a considerable number of nodes or somehow increasing the dimension of covariance matrix will likely run out of memory and, even if that does not happen, the whole processing computation cost is high, including Cholesky decomposition and matrix-vector multiplication. MA involves decomposition in the convolution sense of covariance function, which may influence the applicability regarding computational cost and memory requirements. Although the limitation of TBS had been overcome with computational developments, it was usually associated with the use of few lines to generate random fields, which may introduce artifact effects into them (Emery & Lantuéjoul, 2006). Before each conditional simulation, SGS requires the computation of expected mean and variance. Moreover, SM and LAS are limited to only rectangular grids (simulation mesh), a condition that may be true for some of the above-mentioned methods when simplifications are assumed to relieve the computational processing.

Considering that, the main current techniques applied to geostatistical simulations are Sequential Gaussian Simulation (SGS) (Isaaks, 1990), Turning Bands Simulation (TBS) (Matheron, 1973), and their multivariate versions, Sequential Gaussian Co-simulation (COSGS) (Verly, 1993) and Turning Bands Co-simulation (TBCOSIM) (Emery, 2008). Multivariate simulations or co-simulations are highly recommended for cases with cross-correlated variables, commonly experienced for soil properties.

Some studies have emerged that compare these techniques. Ren (2005) published a short note on conditioning TBS. In this study, the author presented results that demonstrated that the TBS is a fast simulation method when multiple realizations are necessary. For example, performing only one realization of a conditional simulation, SGS was around 5 times faster than TBS. However, increasing this number to 100 realizations, SGS performed around 5.5 times slower than TBS, and this discrepancy presents an almost linear trend in favor of the TBS. Paravarzar et al. (2015) assessed the performance and accuracy of SGS and TBS for jointly simulating co-regionalized variables of a synthetic univariate case and a real multivariate one. The turning bands accurately reproduced the spatial correlation structure for both cases, while the sequential simulation produced some bias, which was more severe in the multivariate case.

The conclusions lead to the claim that TBCOSIM outperforms COSGS in terms of the cross-correlated reproduction, calculated by the spatial continuity and statistical parameters. In addition, the turning bands technique also surpassed other techniques in terms of lower computational costs, standing out more and more when the number of realizations increases. However, studies have not assumed the TBCOSIM technique for applications. Usually, studies have assumed the SGS or other techniques, even assuming fixed functions to describe the spatial correlation of soil properties. Studies have paid little attention to the correct reproduction of the spatial correlation structures of coregionalized variables - or, at least, they do not claim to present the evaluation of their structures. Studies that proposed to investigate the effects regarding spatial variability on probabilistic analysis have focused strictly on the influence of the correlation length (i.e., range) and paid no attention to the agreement between the simulated coregionalization model and the sample one.

Therefore, this paper presents an improved and efficient approach to address probabilistic analysis of geotechnical structures, a geostatistical-based enhancement of the Random Finite Element Method (RFEM) by incorporating an advanced geostatistical technique (i.e., Turning Bands Co-simulation, TBCOSIM), so far not jointly used. It also investigates the influence of correctly reproducing spatial variability on the multivariate probabilistic analysis of geotechnical structure. The primary aim of this work is to provide the correct consideration of the coregionalization model of the soil properties. We illustrate the sophisticated approach and those effects through an actual case of a soil slope, previously presented in the literature.

2. Geostatistical concepts

In reliability studies, geostatistics has gained increasing attention within the area of geotechnical structure design. Geostatistics was originally developed for mining purpose, aiming to characterize the concentrations of certain minerals in a field (Regionalized Variables Theory, RVT) (Matheron, 1973). This theory has two objectives: first, to describe the spatial correlation (theoretically) and, second, to solve estimation problems of a regionalized variable based on a minimal sample (in practice).

The application of a geostatistical technique begins by analyzing the sample data. First, we need to assume that there is a possibility that the value of the random variable for each point, $Z(x)$, in a field is correlated, to some extent, with the values of other nearby points, $Z(x+h)$. This means that the spatial continuity of a regionalized variable can be done with sample values based on two-points statistics. Then, the variogram function, $\gamma(h)$, - used to describe the behavior of spatial correlation in a field - depends only on two points, positioned at a distance h from each other. Analyzing all known data from different points gives the statistical inference for this function. The variogram is calculated as Equation 1.

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}\left[Z(x+h) - Z(x)\right]^2 \tag{1}$$

Where n is the number of pairs of points separated by lag h. Likewise, the cross-variogram function for multivariate fields can be determined as Equation 2.

$$\gamma_{12}(h) = \frac{1}{2n}\sum_{i=1}^{n}\left[Z_1(x) - Z_1(x+h)\right]\cdot\left[Z_2(x) - Z_2(x+h)\right] \tag{2}$$

A theoretical variogram function must then fit the sample covariograms to be applied in simulations. The simple and cross-variograms compose the most important information for geostatistical simulations. However, conventional statistical information (i.e., mean, standard deviation, and probability density function) may also be required.

Determining statistical information requires some precautions, often neglected. For example, site investigations (boreholes) rarely have regular spacing in the field because of difficulty of access, topography, areas of environmental preservation, among other reasons. Thus, sampling may have clustered boreholes, which may introduce bias to statistical inferences. To address this "clustering problem", the declustering technique is recommended (e.g., cell declustering and polygonal declustering methods). In brief, the declustering entails analyzing the "influence area" for each borehole of the campaign and calculates the weights for them (Chilès & Delfiner, 2012).

Once the analysis has determined the variogram function and conventional statistical information of the regionalized or coregionalized variables, geostatistical simulations can be performed. If conditional simulations are desired, it would also require the known data (from the site investigation, boreholes), and their position within the field.

2.1 Turning bands co-simulation - TBCOSIM

TBCOSIM was originally presented by Emery (2008) and developed in MATLAB. It is based on the COSIM program - proposed by Carr & Myers (1985) - and the TBSIM - proposed by Emery & Lantuéjoul (2006). TBCOSIM presents significant improvements compared to previous proposals, which are worth mentioning:

- it allows three-dimension simulations, by grid or scattered points;
- it imposes no restrictions on the number of nested structures, known data points or random variables;
- it works with heterotopic data sets;
- it uses stationary and intrinsic models;
- it uses simple kriging, ordinary kriging or intrinsic co-kriging, associated with the consideration of a unique or moving neighborhood, to condition the simulations to a data set;
- it accepts 15 commonly used covariance models (spherical, exponential, gamma, stable, cubic, Gaussian, cardinal sine, J-Bessel, K-Bessel, generalized Cauchy, exponential sine, linear, power, mixed power and spline), as illustrated in Figure 1;
- it backtransforms variables from Gaussian space to the original units of each variable;
- it can change the support (regularization) of simulations.

Besides adapting and modifying the TBCOSIM for this study, it was also entirely reprogrammed using the Fortran language to achieve the objectives. Next, the code was linked to the RFEM to compose the sophisticated approach used in this study (we will call it "sRFEM"). For further details on the TBCOSIM technique, readers are referred to Emery (2008).

3. Random Finite Element Method - RFEM

Originally, RFEM was proposed by Griffiths & Fenton (1993), and it is considered a powerful and rigorous tool to take the spatial variability of soil properties into account for probabilistic analysis. It uses random field theory (RFT) jointly with the finite element method (FEM).

The FEM is used to compute the plane strain deformation of elastic-perfectly plastic soils governed by the Mohr-Coulomb failure criterion. It is also based on the strength reduction method (SRM) and uses eight-node rectangular quadrilateral elements, with reduced integration (four Gauss points per element) in the generation of the gravity loads, the stiffness matrix and the stress redistribution phase. The adopted solution procedure, to model material non-linearity, is the "constant stiffness" (modified Newton-Raphson) method. For more details, readers are referred to Griffiths & Lane (1999) and Smith et al. (2013).

RFEM uses the Local Average Subdivsion (LAS) method, presented by Fenton & Vanmarcke (1990), to generate the

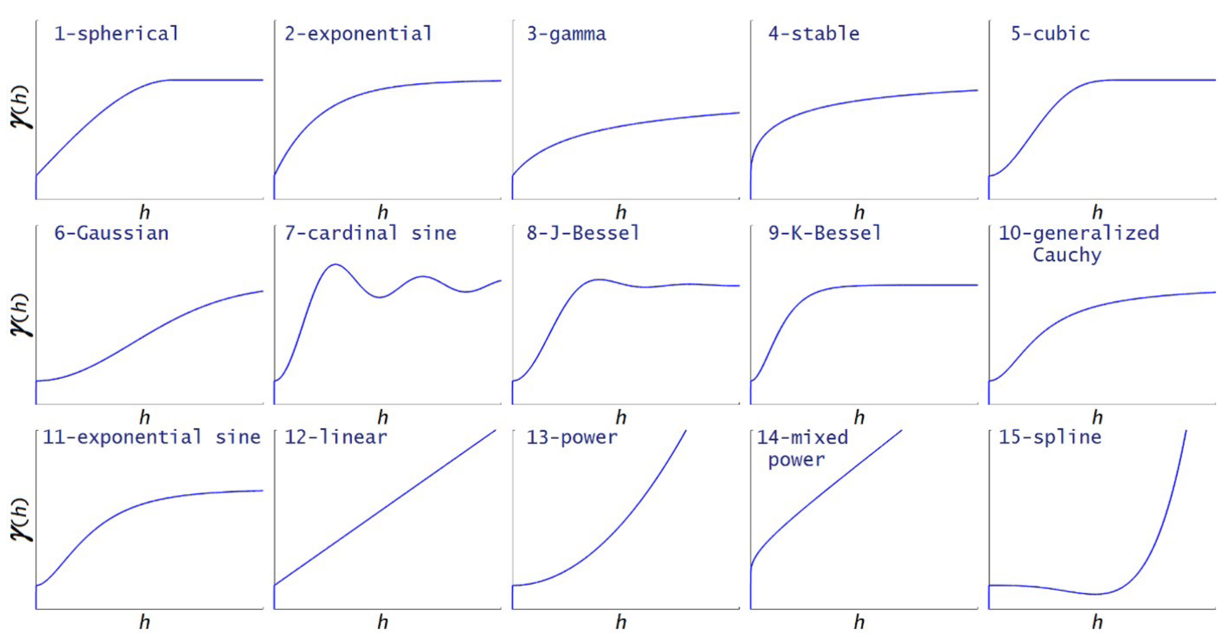

Figure 1. Theoretical fitting models accepted by the TBCOSIM technique.

random fields of simulations. In order to describe the spatial correlation between different spaced points in the field, the LAS method can be used assuming five covariance functions, almost all associated with an exponential decay, and already awarded in the code (RFEM). The most commonly used and recommended function is the Markovian covariance function, which calculates the correlation coefficient (C) between soil properties at different points in the field, as Equation 3.

$$C(h) = exp\left\{-\sqrt{\left(\frac{2h_x}{a_x}\right)^2 + \left(\frac{2h_y}{a_y}\right)^2}\right\} \quad (3)$$

Where a_x and a_y are the spatial correlation ranges in x (horizontal) and y (vertical) directions, respectively. However, for assumed isotropic fields ($a = a_x = a_y$), it can be simplified to Equation 4.

$$C(h) = exp\left\{-\frac{2h}{a}\right\} \quad (4)$$

Then, the RFEM analyzes the geotechnical structure via the FEM application for each simulated random field. For example, in order to test the probabilistic stability of a soil slope, the *mrslope2d* pack executes the strength reduction method. In each simulation, the analysis assumes the failure threshold condition to determine whether it is in the success or failure domains. Finally, the probability of failure (P_f) is obtained by the $(number\ of\ failures)/(number\ of\ realizations)$ ratio. In brief, RFEM uses the Monte Carlo simulation (MCS).

4. Sophisticated RFEM – sRFEM

Besides incorporating the TBCOSIM to enable an advanced geostatistical-based enhancement of the RFEM that, to the best of the authors' knowledge, has not been presented in the literature before, other improvements were also needed and integrated. They were needed to enable the evaluation of specific aspects of the probabilistic analysis that were not presented in the original approach. They are:

- Defining the experimental variograms generated by simulations:
  - o Evaluating the experimental variograms enables the investigation regarding the accordance with and the correct reproduction of the pre-defined coregionalization model (based on the sample one);
  - o Variograms are generated and analyzed in the Gaussian space.
- Defining the factor of safety (*FoS*) calculated for each simulation:
  - o The original source-code evaluates only the limit state condition (*FoS*=1), concluding if the simulated structure is or is not inside the failure domain;
  - o Evaluating the *FoS* for each simulation is done through an iterative process until convergence, considering a pre-defined tolerance value. This process was incorporated into RFEM regarding the probabilistic and deterministic analysis. In the deterministic analysis, the *FoS* used to be calculated by assuming some hypotheses values (low precision), not by an iterative process (convergence);
  - o Storing the *FoS* values, calculated for the structure for each simulation, enables the evaluation of the frequency distribution, or the probability density function (PDF), of the *FoS* of the analyzed structure;
  - o Evaluating the PDF of *FoS* enables revalidation of the calculated results because the variable's (*FoS*) variance can be graphically illustrated, just as the mean value and its behavior (distribution type), close to the peak or in the tailings;
  - o Many programs use the PDF of the *FoS* to estimate the reliability index (β) of the structure.
- Analyzing the convergence of the probability of failure (P_f) with simulations:
  - o The P_f's convergence is an important indicator that should be evaluated when performing a probabilistic analysis because it represents the precision of the calculated value;
  - o According to Melchers & Beck (2017), the MCS requires around $10^{(P+2)}$ simulation to obtain a good estimate of the P_f of a system, where p is the expected order of the P_f of the investigated structure (P_f=x×10^{-P}). Hence, when expecting a significantly low P_f, the total amount of required simulations would be clearly substantially high;
  - o Evaluating the P_f convergence with simulations can show a satisfactory stabilization for a lower or a higher amount than that recommended by Melchers & Beck (2017). Because of that, enabling this evaluation is really important.
- Updating parts of the open source-code with functions and update syntaxes that are more reliable and agile than its precursors:
  - o The functions and codes are constantly updated to attribute more agility and reliability to the programming. Therefore, revisions and updates of previous codes may be needed and recommend;
  - o Since the developed Fortran code, for the application of the TBCOSIM, is a recent programming (developed in this study), updating the available RFEM code ensures better compatibility between the algorithm frameworks.

The following steps summarize the process of performing a conditional probabilistic analysis of a geotechnical structure using the embraced sophisticated approach:

(1) treating and analyzing known data together with their locations in the field, which comprises:

− organizing data in a text file;

- applying the declustering technique for data, which determines weights for each data and borehole (the *declus* algorithm from the Geostatistical Software Library, GSLIB, may be used for this step);
- constructing the histograms considering the declustering weights;
- determining conventional statistical information based on the previous step (mean, standard deviation or coefficient of variance, and type of distribution);
- transforming data from original units into Gaussian space (the *nscore* algorithm from the GSLIB may be used for this step);
- creating a text file with a table to allow the backtransformation of each random variable (needed in later steps);
- calculating the sample variograms (simple and crossed) based on the normalized data (the *gamv* algorithm from the GSLIB may be used for this step);
- fitting the sample variograms by theoretical variograms (using or not nested structures);
- identifying which data will condition the realizations and creating a text file with them.

(2) defining the geometry of the geotechnical structure (a slope in this paper), the mesh dimensions, element size, number of realizations and other parameters for the execution of RSLOPE2D (part of RFEM);
(3) discretizing the structure/field, storing the central position of each element that comprises the mesh;
(4) carrying out a deterministic analysis of the problem;
(5) generating the conditional random fields (one per realization), using the TBCOSIM technique, and storing them

- applying the turning bands technique to generate unconditional random fields;
- using the kriging technique and Equation 5 to condition these fields to the sample data.

$$Z_{cc}(\boldsymbol{x}) = Z_{ci}(\boldsymbol{x}) + \left(Z_{kc}(\boldsymbol{x}) - Z_{ki}(\boldsymbol{x})\right) \quad (5)$$

where $\boldsymbol{x}$ finds the points in space, $Z_{cc}(\boldsymbol{x})$ determines the value in the conditional random field, $Z_{ci}(\boldsymbol{x})$ means the value in the unconditional random field, $Z_{kc}(\boldsymbol{x})$ is the value in the kriging field based on the sample data, and $Z_{ki}(\boldsymbol{x})$ represents the value in the kriging field assuming unconditional data replacing known data.

- backtransforming the simulated values from the Gaussian space to the original units;
- storing the realizations.

(6) analyzing the safety of the structure for each simulated conditional random field;
(7) concluding the probabilistic analysis identifying the probability of failure (P_f) - Monte Carlo simulation (MCS).

All the above-mentioned text files follow the same standardization formats according to the GSLIB's specifications (Deutsch & Journel, 1997).

5. Case study

In order to investigate and to illustrate the effects regarding the reproduction of the spatial correlation structures of soil properties on probabilistic analyses, this paper investigates a real soil slope previously presented in the literature by Johari & Fooladi (2020). According to the reference, the site is in the city of Shiraz, Iran. It has fifteen boreholes, with depths around 25 ~ 26 m from the ground surface. Figure 2 outlines the positioning of these boreholes in the investigated field. Johari & Fooladi (2020) presented all the sample data used in this paper. Readers are referred to this reference for more details.

The analysis assumed a plane strain model. Since the data and the site present a three-dimensional arrangement, the analysis assumed a simulation section. Figure 2 positions the section in the field, while Figure 3 represents it. It is worth mentioning that, for a real design intention, other sections should also be investigated.

For the conditional simulation, only the highlighted boreholes were assumed as conditioning data on its perpendicular projections on the section plane. The conditional

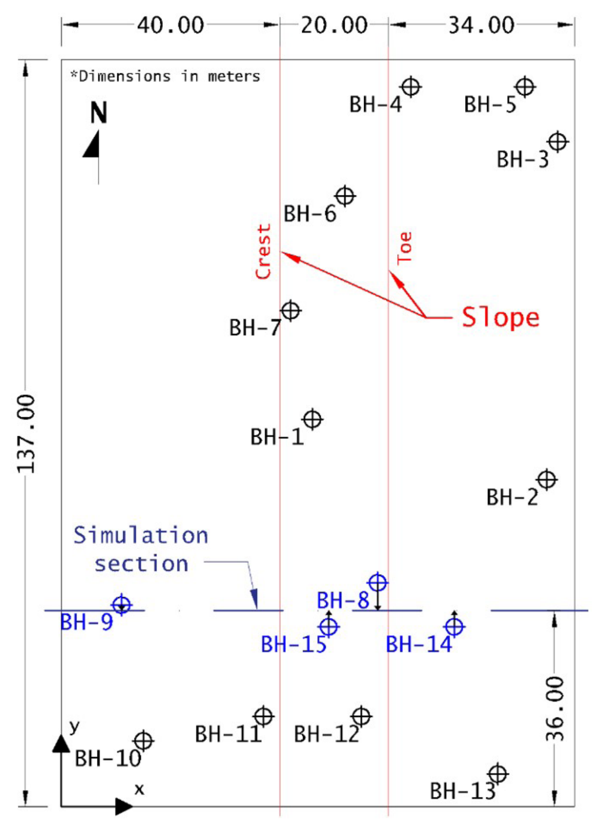

Figure 2. Site representation with borehole locations, slope and assumed simulation section. Modified from Johari & Fooladi (2020).

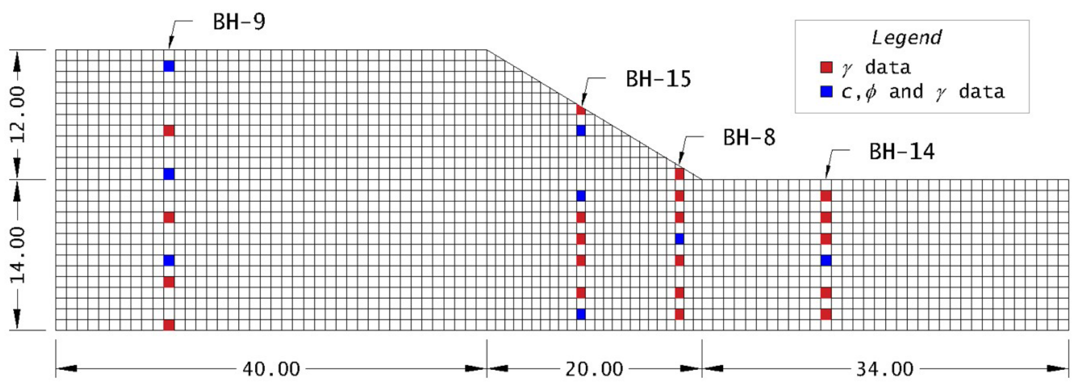

Figure 3. Simulation section with conditioning data locations (distances in meters).

Table 1. Conventional statistical data considering and not considering the declustering.

Soil parameters	Not declustering			Declustering		
	μ	σ	CoV	μ	σ	CoV
c (kPa)	14.89	8.487	0.57	14.02	9.379	0.67
ϕ (º)	25.21	6.050	0.24	25.12	5.782	0.23
γ (kN/m^3)	17.38	0.869	0.05	17.41	0.928	0.05
E and ν were deterministic parameters						
E (kN/m^2)	35,000					
ν	0.30					

μ – mean; σ – standard deviation; ü – coefficient of variation.

simulation did not consider other boreholes because they were distant from this section, differing from the approach assumed by the cited reference. However, all boreholes were used to determine the statistical information on the site parameters.

5.1 Structural analysis

First, since it is an example of an illustration, the field was deemed as an isotropic soil layer. In other words, although the sophisticated approach can deal with this condition, for the sake of simplicity, the anisotropy of the spatial variability was not considered in this paper.

Then, it can be observed that the location of the boreholes, in Figure 2, demonstrates an irregular spatial investigation. As previously mentioned, clustered samples can introduce bias in statistical information. To deal with this condition, the cell declustering technique was performed. For this step, the analysis makes use of the well-known Geostatistical Software Library (GSLIB) (Deutsch & Journel, 1997), specifically the *declus* algorithm.

In agreement with Johari & Fooladi (2020), cohesion (c), friction angle (ϕ) and unit weight (γ) were assumed as random variables, or regionalized variables, while dilatation angle, elastic modulus (E) and Poisson's ratio (ν) were deterministic. Table 1 shows the statistical information presented by the reference (not declustering) and the recalculated one, considering the declustering technique. This shows that the mean and standard deviation values may vary significantly or not when declustering is considered. Although these variations can be low for this illustration, they may lead to different conclusions for the analysis.

The next step of the analysis comprises the generation of the transformation tables (from the original units to the Gaussian space, and vice versa). GSLIB was used once more for this task, specifically the *nscore* algorithm.

Once all the data were in the Gaussian space, they were analyzed to determine the variograms for the field. This determination used all the data from all the boreholes. GSLIB's *gamv* algorithm can calculate these variograms, therefore it was incorporated into the approach. Figure 4 shows the simple and cross-variograms for the sample data.

For the application, a theoretical fitting function to the sample variograms should be determined. To define this function, first it was assumed to be composed of a nugget effect jointly with two spherical nested structures. The spherical spatial correlation function can be expressed as Equation 6.

$$
\begin{aligned}
C(h) &= C\left\{\frac{2}{3}\left(\frac{h}{a}\right)-\frac{1}{2}\left(\frac{h}{a}\right)^3\right\},\ if\ h<a \\
C(h) &= C,\ if\ h>a
\end{aligned}
\tag{6}
$$

Where h is the vector lag between points in the field, C is the sill variogram value, and a is its range or correlation length.

Then, the fitting parameters (of the coregionalization model) were defined thought (1st) a manual fitting, followed by (2nd) applying the weighted least squares method and (3rd) another manual fitting, as refinement, all assuming the requirement of obtaining a licit and positive semi-defined theoretical model. Therefore, the fitted linear function of the coregionalization model was described as Equation 7.

$$C(\boldsymbol{h}) = \begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.48 & 0.26 \\ 0.00 & 0.26 & 0.55 \end{bmatrix} nugget + \begin{bmatrix} 1.00 & -0.48 & -0.57 \\ -0.48 & 0.30 & 0.30 \\ -0.57 & 0.30 & 0.40 \end{bmatrix} sph_{16.0}(\boldsymbol{h}) + \begin{bmatrix} 0.30 & -0.33 & -0.15 \\ -0.33 & 0.40 & 0.13 \\ -0.15 & 0.13 & 0.17 \end{bmatrix} sph_{34.0}(\boldsymbol{h}) \quad (7)$$

Where the first spherical structure persists for 16.0 meters in the range, while the second continues up to 34.0 meters.

5.2 Stochastic analyses

Three assembled configurations were applied to the case study. First, a probabilistic analysis was performed using the original RFEM approach, accordingly with the common seen applications. Second, a new assessment was performed but applying the sRFEM (incorporating the TBCOSIM technique) with no conditional data (unconditional simulation). Finally, the third configuration was similar to the second, but this time the conditioning data of the boreholes close to the simulation section were considered (conditional simulation).

For the conditional simulation, only the highlighted boreholes were assumed as conditioning data, specifically its perpendicular projections on the section plane, see Figure 2 and Figure 3. The conditional simulation did not consider other boreholes because they were distant from this section, differing from the approach assumed by the cited reference. However, all boreholes were used to determine the statistical information.

All configurations assumed the statistical information presented in Table 1, considering the declustering method. In addition, configurations performed 2,000 realizations each, but this amount would increase as needed.

Replicating the reference, the log-normal distribution type describes all the probability density functions (PDFs) of the random variables. However, the correlation length and the correlation matrix were based on the results of the structural analysis, Figure 4. The correlation length was 34.0 meters (range of the variograms), and the correlation matrix (sills of the variograms) was as Equation 8.

$$C(\boldsymbol{h}) = \begin{bmatrix} 1 & -0.81 & -0.72 \\ -0.81 & 1 & 0.43 \\ -0.72 & 0.43 & 1 \end{bmatrix} \quad (8)$$

Conversely, the other configurations used the transformation tables and the theoretical coregionalization function (Equation 7). For the conditional simulation, third configuration, the ordinary cokriging technique was selected to condition the simulated fields to the known data.

5.3 Spatial covariance reproduction

The first question that arises from the application of a geostatistical simulation is whether it complies with the reproduction of the spatial covariance model. A simple procedure to assess this condition is to analyze the experimental variograms generated by the simulations and compare them with the input model.

Figure 5 presents this evaluation for the second configuration (unconditional simulation using the sRFEM). Note that for each realization, the experimental variograms change, moving away from or closer to the theoretical model (cloud of simulation). However, when analyzing the average experimental variograms, they should agree with the theoretical coregionalization model. Therefore, the unconditional simulation via sRFEM satisfactorily reproduces the spatial covariance model as the average variograms, both simple and crossed, agree with the input model.

Figure 6 shows a comparison of the average variograms for all configurations and the theoretical coregionalization model. The first configuration, using the original RFEM, could not correctly reproduce the model of spatial covariance. Note that the generated variograms resulted in a lower covariance between the equispaced points compared with the theoretical model. Then, this condition can implicate in a random field with a lower variance when compared to the real one, leading to a non-conservative analysis. Despite not adhering to the theoretical structure behavior, the major contribution to the observed discrepancy is related to the nugget effect, which is ignored in this configuration (null nugget effect, with variograms starting from origin).

Otherwise, as previously mentioned, the unconditional simulation via sRFEM successfully reproduced the theoretical model. However, when analyzing the conditional simulation, note that a disturbance occurs for simple variograms. The agreement is hardly supported by conditional simulations. There are some reasons for this “disagreement”. The primary reason is that the conditioning data does not perfectly fit the theoretical model assumed for the simulation, as can be seen in Figure 4. In other words, the known data in the field introduce a certain “distortion” regarding the input or fitted model. Despite this effect, the cross-variograms for the conditional simulation were satisfactorily in accordance with the prior model.

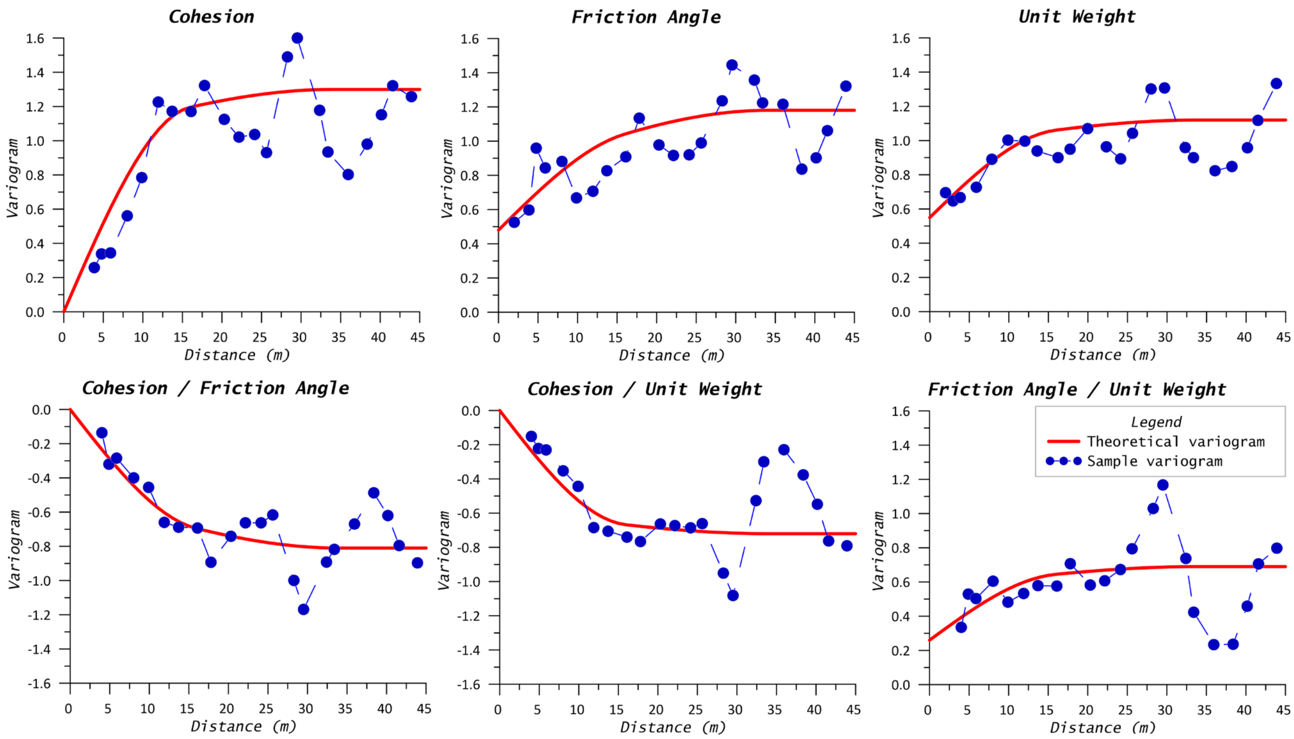

Figure 4. Fitting of sample variograms by theoretical nested structures - simple variograms (top) and cross-variograms (down).

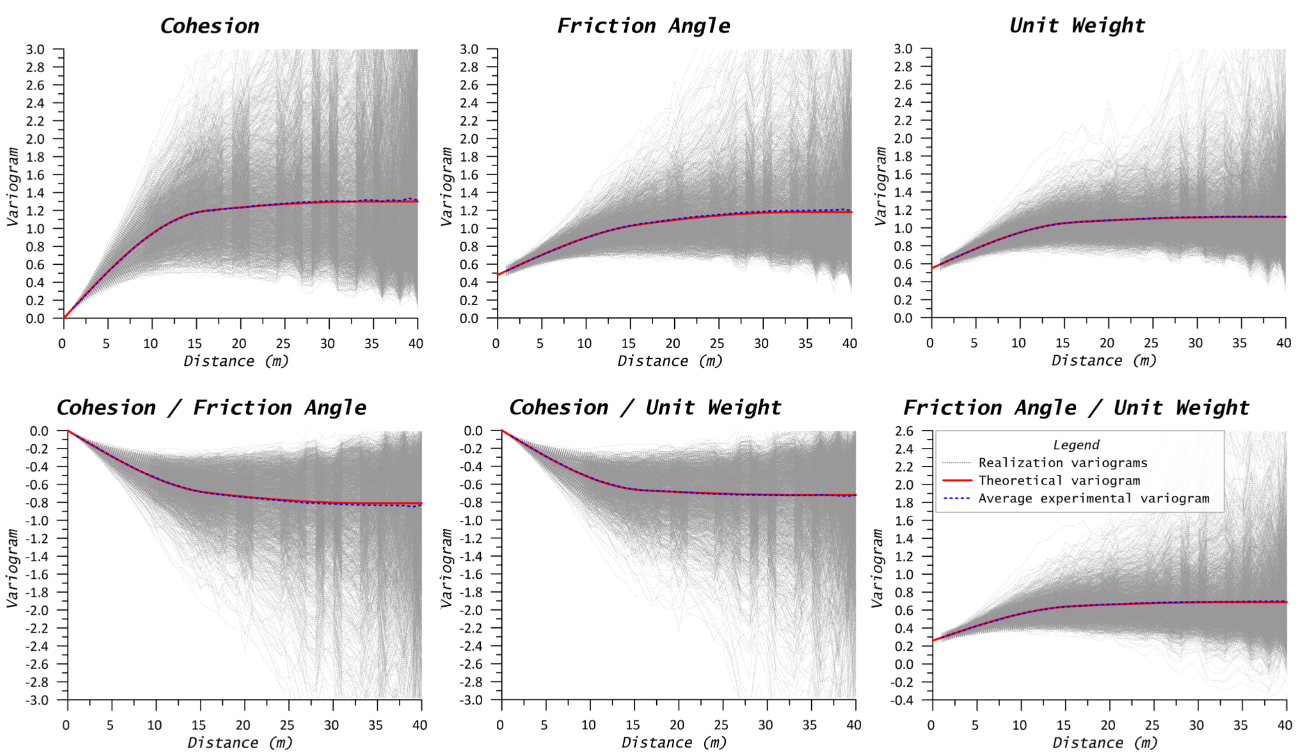

Figure 5. Comparison between the average experimental and the theoretical variograms for the unconditional simulation.

A section simulated example is shown in Figure 7. It illustrates the first realization of the conditional random field for the case study. In addition, the displaced FE mesh was jointly presented.

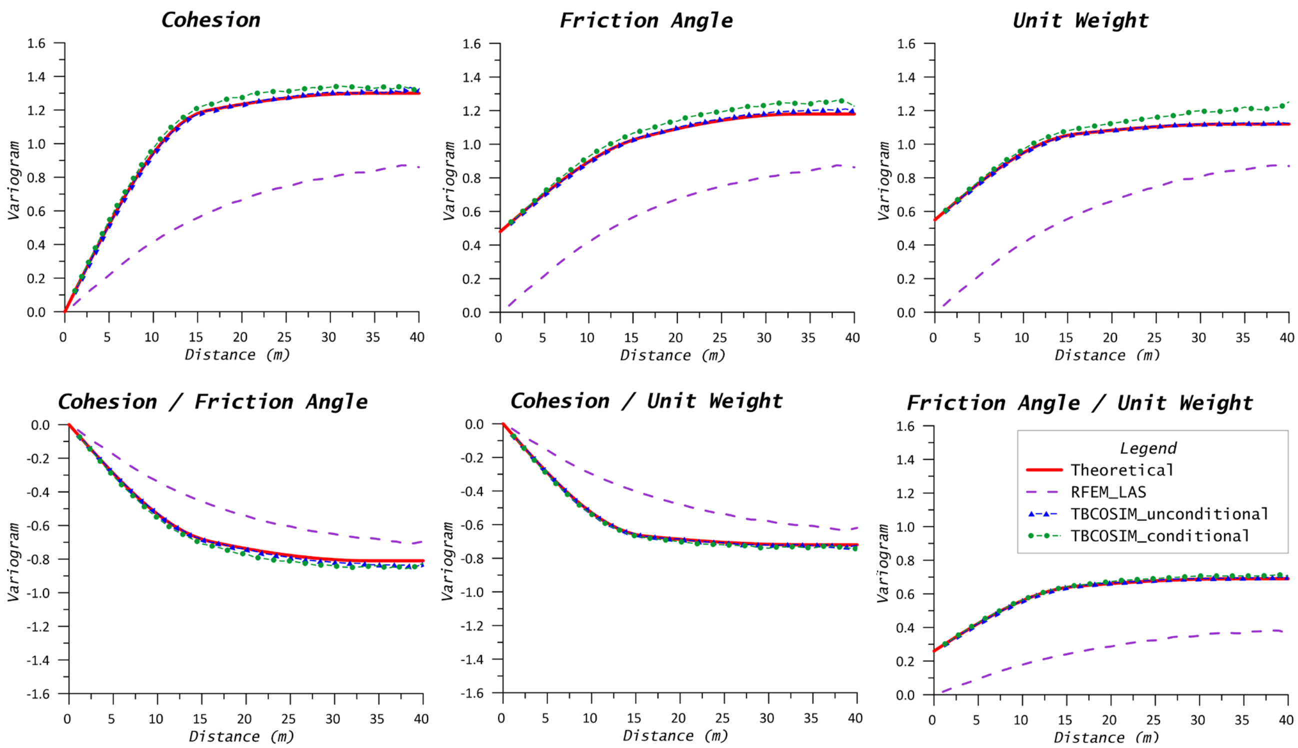

Figure 6. Comparison between the average experimental and theoretical variograms using three methods – Original RFEM, unconditional and conditional using sRFEM.

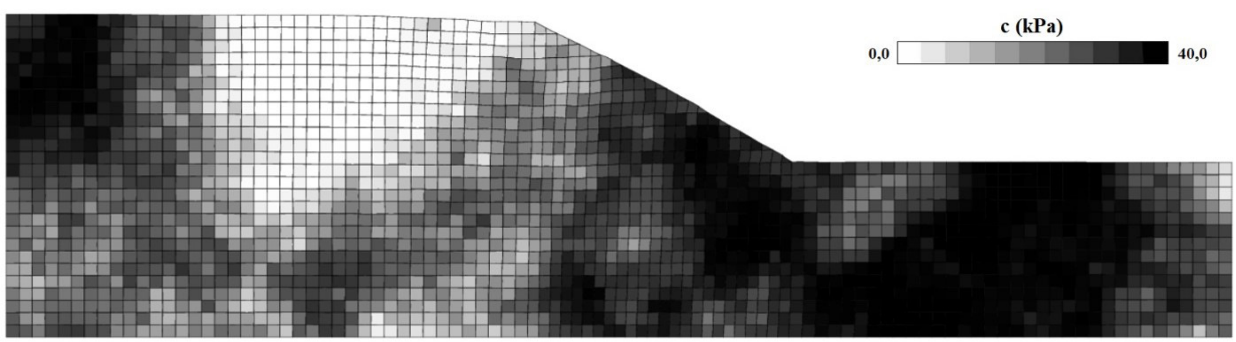

Figure 7. Simulation section of a conditional realization using the sRFEM, representing the displaced FE mesh and the random field in terms of the cohesion parameter.

5.4 Failure assessment

After performing the simulations, the MCS technique was jointly used to define the probability of failure (P_f) of the geotechnical structure. The stability analysis was carried out using the approach incorporated by the RFEM (the strength reduction method), with some adaptations, such as mentioned in the item 4. As the stability analyses were performed, the P_f was monitored, allowing the investigation of its convergence, as shown in Figure 8. A failure event is defined to occur when the factor of safety (FoS) is less than the unit ($FoS < 1$).

Storing the FoS for each realization also grants the investigation of its ensemble (PDF), as shown in Figure 8. Note that both the convergence of the P_f and the ensemble for the FoS were analyzed for the three configurations.

A deterministic stability analysis of this structure, using only the mean values presented in Table 1 (with declustering), resulted in the ü equal to 1.56. In agreement with this deterministic result, the PDFs for FoS of the three configurations have the peak of their distributions around this value, and very close to the value presented by Johari & Fooladi (2020). However, the distributions showed different behaviors between them, mainly regarding the variance and the behavior of the upper

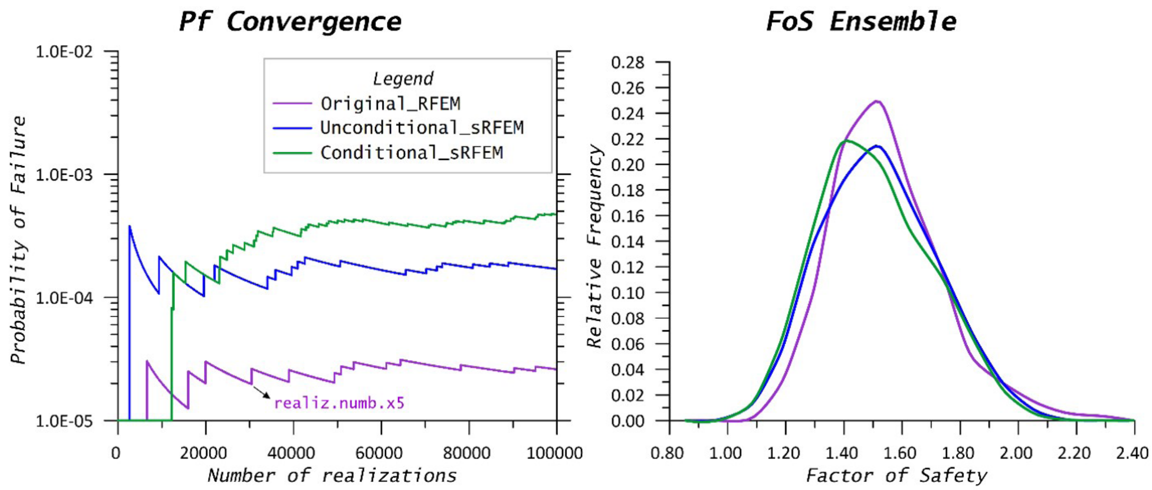

Figure 8. Convergence of the probability of failure and PDF of the factor of safety (FoS) through each method.

and lower tails. The first approach (using the original RFEM) led to a lower variance compared to both unconditional and conditional analysis using the sRFEM. In addition, the P_f for first approach was lower than the second one (around 0.003% and 0.017%, respectively), as expected when examining their distributions. The P_f for the first configuration was obtained from an MCS with 5×10^5 realizations, because of its lower value, while 10^5 simulations were assumed for the second and the third approaches.

In contrast to previous work exploring conditional simulations for this purpose, the conditional simulation (third configuration, using the sRFEM) led to a higher P_f compared with previous configurations (around 0.047%). A PDF analysis reaffirms this result, because the lower tail distribution for the conditional simulation presents a slightly larger area (for $FoS < 1$) than the unconditional one. In addition, the conditional approach also showed a marginally higher variance than the unconditional simulation. Usually, these results are unexpected because conditional simulations are used to reduce the uncertainties about the field variance. However, in this case study, this "unanticipated" condition can occur for a few reasons:

- as previously mentioned, conditional data are used to introduce a disturbance to the coregionalization model, which can be seen in Figure 6 mainly for the simple variograms. Then, note that this effect, in this case, shifted the simple variograms to higher values, so for the same lag vector between any two points, the variance is higher for the conditional than for the unconditional simulation;
- since the geostatistical parameters were defined based on all available data (three-dimensional field), the data in the simulated section may have a slight discrepancy regarding the generalized covariance model, e.g., conditioning data presents higher variance than the entire set of data regarding the investigated field;
- the behavior close to the origin of the spatial covariance models has a substantial influence in the final simulated variances. High nugget effects lead to higher variances in results. If the definition of this behavior is not well-founded, it may produce loss of efficiency of the kriging techniques, and the computed variance may be inaccurate. This effect has even more serious consequences for the conditional simulations, which aim to reproduce the variability in greater detail (Chilès & Delfiner, 2012);
- all the conditioning data assumed for the simulated section (boreholes 8, 9, 14 and 15) have values for the friction angle parameter lower than the mean value for the entire investigated field. Then, it can lead the conditional strength to a reduced average in lower tail realizations;
- perhaps, the number of investigations (boreholes) or points with known data was not enough to reduce the uncertainty level, as explored and presented by Yang et al. (2019).

In the meantime, it is worth mentioning that the PDF, for the conditional simulation, around the peak (close to the mean value of FoS) has a bottleneck format. It suggests that realizations around the mean values granted a lower variance than the peripheral ones.

6. Conclusion

This paper addressed the correct reproduction of the spatial coregionalization model and investigated the effects regarding this reproduction for probabilistic analysis of geotechnical structures. For this, it uses geostatistical concepts and advanced techniques (TBCOSIM) jointly with the most reliable and applied approach presented in the literature to deal with random fields associated with the FEM (RFEM). Based on the results for the illustrative case study, the following conclusions can be drawn:

- Determining the simple and cross-variograms, for the sample data and the fitting theoretical coregionalization function, is an important task in geostatistical treatments, hence also in a probabilistic analysis of geotechnical structures;
- The spatial covariance reproduction when using the sRFEM satisfactorily agrees with the input coregionalization model. The average variograms of the unconditional simulation almost perfectly agree with the theoretical ones, while the conditional one presents a small shifting factor of the variograms upwards (higher values of variance), since the known data rarely agree precisely with the "fitting" model. Otherwise, the original RFEM, as a common approach, failed in this reproduction, leading to lower variances than the sRFEM. Therefore, the RFEM would present a non-conservative design for this structure, resulting in an expressive low P_f, which may be a consequence of the failure in reproducing the spatial variability, for this case;
- Neglecting the investigation distribution in the field may lead to bias statistical information about it. The declustering technique is an important tool to deal with clustered investigations, often seen in practice;
- Disregarding the nugget effect, simulations cannot characterize the local uncertainty (e.g., uncertainties regarding measurements, equipment, tests, correlation formulas, and other sources). It affects the reproduction of the coregionalization model, hence may lead the analysis to biased results;
- Although previous studies often associating the conditional simulation with lower variances and probability of failure, this "expected" condition was not observed for this case study. This is because of some factors presented at the end of Section 5.4;
- Depending on amount and location of the conditioning data, jointly with the geostatistical structures, conditional simulations may not offer a meaningful reduction in the simulation's variance. However, the result may be considered more reliable than before, e.g., in the study case, since the values of the conditioning strength parameters around the section were lower than the mean value for the field (entire set), the peak value of *FoS* for the conditional simulation became lower than the peak value for the unconditional one;
- Incorporating the TBCOSIM into the RFEM produces an improved and efficient approach to deal with probabilistic analysis of geotechnical structures, complying with the spatial correlation structures of soil properties, which comprise them.

Acknowledgements

The authors would like to acknowledge the financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

Declaration of interest

The authors have no conflicts of interest to declare. All co-authors have observed and affirmed the contents of the paper and there is no financial interest to report.

Authors' contributions

Jean Lucas dos Passos Belo: conceptualization, data curation, methodology, validation, writing – original draft. Paulo Ivo Braga de Queiroz: supervision, methodology, validation, writing – review & editing. Jefferson Lins da Silva: supervision, writing – review & editing.

List of symbols

a	Spatial correlation range in the field
a_x	Horizontal correlation range in the field
a_y	Vertical correlation range in the field
BH	Borehole
c	Cohesion
C	Sill of the variogram
$C(h)$	Correlation function
COSGS	Sequential Gaussian Co-simulation
CoV	Coefficient of variation
CRF	Conditional Random Field
E	Young's modulus
FE	Finite element
FEM	Finite Element Method
FoS	Factor of safety
GSLIB	Geostatistical Software Library
h	Spatial lag between two points in the field
h_x	Horizontal lag between two points in the field
h_y	Vertical lag between two points in the field
KLE	Karhunen-Loève Expasion
LAS	Local Average Subdivision
LEM	Limit Equilibrium Method
MA	Moving Average Method
MCS	Monte Carlo simulation
MD	Matrix Decomposition Method
N	North
p	Expected order for the probability of failure
PDF	Probability Density Function
P_f	Probability of failure
RFEM	Random Finite Element Method
RFT	Random Field Theory
RVT	Regionalized Variables Theory
SGS	Sequential Gaussian Simulation

SM	Spectral Method
sph	Spherical function
sRFEM	Sophisticated Random Finite Element Method
SRM	Strength Reduction Method
TBCOSIM	Turning Bands Co-simulation
TBS	Turning Bands Simulation
x	Point in the field
Z(x)	Value of the random variable for a point in the field
Z_{cc}	Value of the random variable in the conditional random field
Z_{ci}	Value of the random variable in the unconditional random field
Z_{kc}	Value of the random variable in the kriging field based on the sample data
Z_{ki}	Value of the random variable in the kriging field based on the simulated unconditional data replacing known data
β	Reliability index
ϕ	Friction angle
γ	Unit weight
γ(h)	Simple variogram function
$\gamma_{12}(h)$	Cross-variogram function
μ	Mean
ν	Poisson's ratio
σ	Standard deviation

References

Belo, J.L.P., & Silva, J.L. (2020). Reliability analysis of a controlled stage-constructed and reinforced embankment on soft ground using 2d and 3d models. *Frontiers in Built Environment*, 5(150), http://dx.doi.org/10.3389/fbuil.2019.00150.

Belo, J.L.P., Silva, J.L., & Queiroz, P.I.B. (2022). Reliability analysis of an embankment built and reinforced on soft ground using LE and FE. *Sustainability*, 14(4), 2224. http://dx.doi.org/10.3390/su14042224.

Carr, J.R., & Myers, D.E. (1985). Cosim: a fortran iv program for coconditional simulation. *Computers & Geosciences*, 11, 675-705. http://dx.doi.org/10.1016/4830098-3004(85)90012-3.

Chilès, J., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118136188.

Deutsch, C.V., & Journel, A.G. (1997). *GSLIB: geostatistical software library and user's guide.* (2nd ed.). Oxford University Press.

Emery, X. (2008). A turning bands program for conditional co-simulation of cross-correlated gaussian random fields. *Computers & Geosciences*, 34, 1850-1862. http://dx.doi.org/10.1016/j.cageo.2007.10.007.

Emery, X., & Lantuéjoul, C. (2006). Tbsim: a computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences*, 32, 1615-1628. http://dx.doi.org/10.1016/j.cageo.2006.03.001.

Fenton, G.A., & Vanmarcke, E.H. (1990). Simulation of random fields via local average subdivision. *Journal of Engineering Mechanics*, 116, 1733-1749. http://dx.doi.org/10.1061/(ASCE)0733-9399(1990)116:8(1733).

Folle, D., Costa, J., Koppe, J., & Raspa, G. (2006). Metodologies for soil heterogeneities quantification for use on geotechnical engineering. *Solos e Rochas*, 29(3), 297-310.

Griffiths, D.V., & Fenton, G.A. (1993). Seepage beneath water retaining structures founded on spatially random soil. *Geotechnique*, 43, 577-587. http://dx.doi.org/10.1680/geot.1993.43.4.577.

Griffiths, D.V., & Lane, P.A. (1999). Slope stability analysis by finite elements. *Geotechnique*, 49, 387-403. http://dx.doi.org/10.1680/geot.1999.49.3.387.

Griffiths, D.V., Huang, J., & Fenton, G.A. (2009). Influence of spatial variability on slope reliability using 2-d random fields. *Journal of Geotechnical and Geoenvironmental Engineering*, 135(10), 1367-1378.

Isaaks, E. (1990). *The application of Monte Carlo methods to the analysis of spatially correlated data* [Unpublished doctoral dissertation]. Stanford University.

Johari, A., & Fooladi, H. (2020). Comparative study of stochastic slope stability analysis based on conditional and unconditional random field. *Computers and Geotechnics*, 125, 103707. http://dx.doi.org/10.1016/j.compgeo.2020.103707.

Johari, A., & Gholampour, A. (2018). A practical approach for reliability analysis of unsaturated slope by conditional random finite element method. *Computers and Geotechnics*, 102, 79-91. http://dx.doi.org/10.1016/j.compgeo.2018.06.004.

Jurado, C.S., Breul, P., Bacconnet, C., & Benz-Navarrete, M. (2020). Probabilistic 3d modelling of shallow soil spatial variability using dynamic cone penetrometer results and a geostatistical method. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 15(2), 139-151. http://dx.doi.org/10.1080/17499518.2020.1728558.

Kim, J.M., & Sitar, N. (2013). Reliability approach to slope stability analysis with spatially correlated soil properties. *Soil and Foundation*, 53, 1-10. http://dx.doi.org/10.1016/j.sandf.2012.12.001.

Li, Y., Hicks, M., & Vardon, P. (2016). Uncertainty reduction and sampling efficiency in slope designs using 3d conditional random fields. *Computers and Geotechnics*, 79, 159-172. http://dx.doi.org/10.1016/j.compgeo.2016.05.027.

Liu, L.-L., Cheng, Y.-M., & Zhang, S.-H. (2017). Conditional random field reliability analysis of a cohesion-frictional slope. *Computers and Geotechnics*, 82, 173-186. http://dx.doi.org/10.1016/j.compgeo.2016.10.014.

Liu, Y., Li, J., Sun, S., & Yu, B. (2019). Advances in Gaussian random field generation: a review. *Computational*

Geosciences, 23, 1011-1047. http://dx.doi.org/10.1007/s10596-019-09867-y.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5, 439-468. http://dx.doi.org/10.2307/1425829.

Melchers, R.E., & Beck, A.T. (2017). *structural reliability analysis and prediction*. (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781119266105.

Monteiro, S.T., Ramos, F., & Hatherly, P. (2009). Conditional random fields for rock characterization using drill measurements. In *2009 International Conference on Machine Learning and Applications*, Miami, December 2009 (pp. 366–371). IEEE. https://doi.org/10.1109/ICMLA.2009.80.

Mrabet, Z., & Bouayed, A. (2000). Reducing uncertainty on the results of reliability analysis of earth fills using stochastic estimations. In *Second International Conference on Computer Simulation in Risk Analysis and Hazard Mitigation*, Bologna, Italy (pp. 203-214). WIT Press.

Muñoz, E., Cordão Neto, M.P., & Ochoa, A. (2018). Effects of spatial variability on slope reliability: a hypothetical case study. *Soils and Rocks*, 41(3), 381-390. http://dx.doi.org/10.28927/SR.4113381.

Paravarzar, S., Emery, X., & Madani, N. (2015). Comparing sequential gaussian and turning bands algorithms for cosimulating grades in multi-element deposits. *Comptes Rendus Geoscience*, 347, 84-93. http://dx.doi.org/10.1016/j.crte.2015.05.008.

Ren, W. (2005). *Short note on conditioning turning bands realization*. Centre for Computational Geostatistics, University of Alberta.

Sayão, A.S.F.J., Sandroni, S.S., Fontoura, S.A.B., & Ribeiro, R.C.H. (2012). Considerations on the probability of failure of mine slopes. *Soils and Rocks*, 35(1), 31-37.

Schöbi, R., & Sudret, B. (2015). Application of conditional random fields and sparse polynomial chaos expansions to geotechnical problems. In *Proceedings of the 12th International Conference on Structural Safety and Reliability* (pp. 1356-1363). Amsterdã: IOS Press.

Smith, I.M., Griffiths, D.V., & Margetts, L. (2013). *Programming the finite element method*. (5th ed.). John Wiley & Sons.

Stefanou, G., & Papadrakakis, M. (2007). Assessment of spectral representation and Karhunen-Loève expansion methods for the simulation of Gaussian stochastic fields. *Computer Methods in Applied Mechanics and Engineering*, 196, 2465-2477. http://dx.doi.org/10.1016/j.cma.2007.01.009.

Sudret, B., & Der Kiureghian, A. (2000). *Stochastic finite element methods and reliability: a state-of-the-art report* (Rep. No. UCB/SEMM-2000/08). University of California at Berkley.

Verly, G.W. (1993). Sequential gaussian cosimulation: a simulation method integrating several types of information. In A. Soares (Ed.), *Geostatistics Tróia '92* (Vol. 1, pp. 543–554). Springer. https://doi.org/10.1007/978-94-011-1739-5_42.

Yang, R., Huang, J., Griffiths, D., & Sheng, D. (2017). Probabilistic stability analysis of slopes by conditional random fields. In J. Huang, G. A. Fenton, L. M. Zhang & D. V. Griffiths (Eds.), *Geo-Risk 2017* (pp. 450-459). ASCE. https://doi.org/10.1061/9780784480717.043.

Yang, R., Huang, J., Griffiths, D., Meng, J., & Fenton, G.A. (2019). Optimal geotechnical site investigations for slope design. *Computers and Geotechnics*, 114, 103111. http://dx.doi.org/10.1016/j.compgeo.2019.103111.