



Time-dependent reliability analysis of unsaturated slopes under rapid drawdown with intelligent surrogate models

Carlotta Guardiani¹ · Enrico Soranzo¹ · Wei Wu¹

Received: 12 April 2021 / Accepted: 19 September 2021 / Published online: 19 October 2021
© The Author(s) 2021

Abstract

Slope stability in reservoirs depends on time-dependent triggering factors such as fluctuations of the groundwater level and precipitation. This paper assesses the stability of reservoir slopes over time, accounting for the uncertainty of the shear strength and hydraulic parameters. An intelligent surrogate model has been developed to reduce the computational effort. The capability of two machine learning algorithms, namely Support Vector Regression and Extreme Gradient Boosting, is considered to obtain the relationship between geomechanical parameters and the factor of safety. The probability of failure of a hypothetical reservoir slope is estimated employing Monte Carlo simulations for different scenarios of drawdown velocity. A sensitivity analysis is conducted to investigate the influence of the geomechanical parameters, regarded as random variables, on the probability of failure. The results revealed that the coefficient of variation in the effective friction angle and the correlation between effective cohesion and friction angle have the highest impact on the probability of failure. The intelligent surrogate model can predict the factor of safety of reservoir slopes under rapid drawdown with high accuracy and enhanced computational efficiency.

Keywords Machine learning · Rapid drawdown · Reliability · Slope stability · Surrogate model · Unsaturated

1 Introduction

Rapid drawdown is one of the major causes of failure of earth slopes subjected to changing river levels. The decrease in the water level reduces the stabilizing effect of the hydrostatic pressure on the slope. With rapid drawdown, the internal pore water pressure (PWP) distribution of a fine-grained slope reflects the initial water level for some time. Three approaches have been proposed to investigate this scenario: undrained analyses, flow methods and coupled flow deformation analyses. The first group is suitable for impervious materials and neglects the water flow since the dissipation of PWP is much slower than the decrease in the water level. Morgenstern [34] adopted this approach and provided stability charts to estimate the factor of safety during rapid drawdown. Lane and Griffiths

[28] developed a chart-based strategy to achieve safe drawdown with 2D Finite Element (FE) undrained analysis. The second approach, which is appropriate for permeable materials, solves the flow problem caused by the change of hydraulic boundary conditions with time. The underlying assumption is that the solid skeleton is rigid, i.e., soil deformations are neglected. A fully coupled flow deformation analysis was first proposed by Pinyol et al. [42]. The authors highlighted the discrepancies between the aforementioned approaches: the undrained analysis leads to conservative and unrealistic results while pure flow analyses underestimate the PWP. The authors validated the coupled hydro-mechanical method by comparing the calculated PWP with piezometer measurements from the Canelles landslide in Spain [43].

These methods address rapid drawdown with a deterministic approach, which uses the definition of the factor of safety (FOS) to assess slope stability. In the last 20 years, there has been growing interest in probabilistic methods for geotechnical engineering, which incorporate the inherent uncertainties of soil properties in the analysis and design of geotechnical structures. In broad terms, reliability-based

✉ Carlotta Guardiani
carlotta.guardiani@boku.ac.at

¹ Institute of Geotechnical Engineering, University of Natural Resources and Life Sciences, Feistmantelstraße 4, 1180 Vienna, Austria

design (RBD) approaches describe the performance of a system as a function of stochastic variables. RBD methods can be classified based on different levels of sophistication: semi-probabilistic (Level I), approximate probabilistic methods (Level II) and fully probabilistic methods (Level III). The first level is already implemented in Eurocode 0 [5] with the introduction of partial safety factors, which are calibrated so that the required probability that a limit state will not be exceeded is satisfied. Levels II and III estimate the probability of failure (p_f) with analytical and numerical methods, respectively.

The first attempts to assess slope stability within a probabilistic framework [1, 7, 31] are based on the first and second-order reliability method (FORM and SORM, Level II) combined with limit equilibrium methods (LEM). The limit state function is defined as $G(\mathbf{x}) = \text{FOS}(\mathbf{x}) - 1$, where \mathbf{x} represents the stochastic variables. In FORM, the limit state function is approximated by the first-order Taylor series expansion (in SORM the Taylor series is expanded to the second term) and the reliability is measured, according to the Hasofer–Lind method, by the minimum distance between the origin of a standardized coordinate system (e.g., $X'_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}$) and the limit state $G(\mathbf{x}) = 0$. The probability of failure can be obtained as $p_f = \Phi(-\beta)$. The main drawback of these methods lies in the approximate estimation of p_f . Alternatively, fully probabilistic methods (Level III) can be employed for an accurate estimation of p_f , but at the expense of high computational effort. For instance, a direct Monte Carlo simulation requires approximately one million trials to estimate a value of p_f equal to 10^{-4} (which most closely corresponds to the target reliability of 3.8 in a 50 years reference period prescribed in Eurocode 0) with a coefficient of variation (COV) of 10% [15]. Conducting this number of numerical simulations is time-prohibitive for complex geotechnical analyses. A few studies have attempted to perform reliability analysis to real case studies of reservoir slopes with Monte Carlo simulations [30, 53, 56]. The estimation of geomechanical parameters in terms of statistical moments and distribution certainly represents one of the greatest challenges in reliability analysis. The other limitation is the computational cost. Two strategies are available to increase the performance of Monte Carlo simulations. The first reduces the sample size with sampling techniques such as latin hypercube sampling (LHS) [48], Importance Sampling and adaptive simulation techniques, e.g., subset simulation [19, 38]. A second way to achieve higher computational efficiency is to approximate the limit state function with a surrogate model or response surface method (RSM). A simple mathematical model or a more complex soft computing algorithm can be applied to ascertain the relationship between the stochastic

parameters and the output of the limit state function or the FOS. The following techniques have been extensively explored in geotechnical reliability analysis: Hermite polynomial chaos expansion (HPCE) [20–22], Kriging-based RSM [55], support vector machines (SVM) [23, 25, 26], Gaussian processes (GP) [24], extreme gradient boosting (XGBoost) [52], artificial neural networks (ANN) [49] and multivariate adaptive regression splines (MARS) [54].

Recently, RBD methods are increasingly applied to slope stability problems due to the augmented computational efficiency that can be reached with surrogate models. Jiang et al. [20] proposed a non-intrusive stochastic FE method based on HPCE to perform slope stability analysis in spatially variable soils. The authors applied this method to estimate the probability of failure of two illustrative examples, namely an undrained clay slope and a $c - \phi$ slope. SVM was successfully employed together with Monte Carlo simulations to solve slope reliability analysis in the following studies [25, 29]. Optimization methods like particle swarm optimization (PSO) and artificial bee colony (ABC) have been integrated with SVM to enhance the accuracy of their predictions [23, 26]. Gao et al. [10] performed a reliability analysis of a slope under rapid drawdown with FORM. Based on the numerical results obtained with the finite difference method (FDM), the authors applied a polynomial regression model to estimate FOS from the following input parameters: slope angle, reservoir water level, hydraulic conductivity and drawdown rate. The last two parameters were considered as stochastic and the impact of their uncertainty on the reliability was investigated. A sensitivity analysis proved that the hydraulic conductivity affects slope reliability with the highest impact. Wang et al. [51] proposed a surrogate model based on MARS to evaluate the probability of failure of an earth dam under transient seepage and compared different scenarios of rising upstream water level velocity. Jiang et al. [22] conducted a reliability analysis of an embankment under unsaturated seepage considering the spatial variability of shear strength and hydraulic properties. The authors developed a surrogate model based on the HPCE and conducted a sensitivity analysis to study the impact of the variability of each random variable on p_f . Wang et al. [52] used the extreme gradient boosting (XGBoost) algorithm to perform reliability analysis of an earth dam. Both machine learning approaches for regression and classification were explored to assess slope stability and estimate p_f .

In this study, an intelligent surrogate model is constructed to estimate the probability of failure of a reservoir slope over time under rapid drawdown with the random variable method. The uncertainties of the shear strength and hydraulic parameters, including the curve fitting

parameters of the soil-water characteristic curve (SWCC), are modeled with a lognormal distribution. The surrogate model is trained on an initial sample of data generated with LHS. The problem is studied in the time frame after rapid drawdown and the surrogate model is fed with the results obtained at different time steps of the analysis. The performance of two machine learning models, i.e., SVM and XGBoost, is evaluated and compared. The surrogate model with the best performance in terms of prediction accuracy and computational efficiency is selected to estimate p_f with time. Section 2 presents the methodology used, namely the deterministic model, the sampling scheme and the surrogate models that are tested. Section 3 presents the results obtained for different scenarios of rapid drawdown. Section 4 includes a sensitivity analysis of the stochastic parameters.

2 Methodology

First, the problem is studied with a deterministic approach, which consists of an unsaturated transient seepage analysis and a slope stability limit equilibrium analysis based on the Morgenstern–Price method. The modules SEEP/W and SLOPE/W available in the commercial software GeoStudio [32] are used to conduct the analyses. A homogeneous slope with a simplified geometry is considered. The slope is 10 m high and has a slope ratio of 2:1 (Fig. 1). The initial condition is a steady state seepage with the reservoir level at 9 m above the foundation (19 m of elevation in Fig. 1). Then, a constant total head of 19 m is applied on the right side of the slope, while a linear function of the total head with time is applied on the left side of the slope (including the horizontal boundary of the foundation). This function represents the linear decrease in the reservoir level with time from 19 to 10 m, where the slope of the function is the drawdown rate indicated for each scenario. A mesh refinement analysis has been conducted to select a suitable mesh size. The results of deterministic simulations with the mean values of the soil parameters provided in

Table 1 Input soil parameters of the slope and the foundation

Analysis	Parameter	μ	COV	Correlation (ρ)
Stability	c'	5 kPa	0.30	$\rho_{c',\phi'} = -0.5$
	ϕ'	30°	0.20	
	γ	20 kN m ⁻³	-	-
Seepage	k_s	1.16×10^{-6} ms ⁻¹	0.60	-
	a	28.44 kPa	0.40	$\rho_{a,n} = -0.25$
	n	1.72	0.20	
	θ_s	0.5 m ³ m ⁻³	-	-
	θ_r	0.035 m ³ m ⁻³	-	-

Table 1 and with increasing mesh size have been compared. It was found that a mesh size of one meter is appropriate for the problem considered in this study. Table 1 shows the soil properties of the slope, which are the same of the foundation.

2.1 Transient seepage analysis

In the seepage analysis, only the soil above the foundation (above 10 m elevation) is considered partially saturated, while for the underlying part only a fully saturated analysis is performed. The governing equation for two-dimensional seepage through unsaturated soil was derived by Richards [44], who extended Darcy's law to unsaturated flow, and can be expressed as:

$$\frac{\partial}{\partial x} \left(k_x \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial h}{\partial y} \right) + Q = \frac{\partial \theta}{\partial t} \quad (1)$$

where k_x and k_y are the horizontal and vertical hydraulic conductivity (i.e., in the x and y direction, respectively), h is the total head, Q is an external applied boundary flux and θ is the volumetric water content. In unsaturated soils, θ varies depending on the matric suction $\psi = u_a - u_w$, which is the difference between the air and the water pressure. This relationship is described by the soil–water characteristic curve (SWCC). In this study, the model proposed by Van Van Genuchten [11] has been adopted:

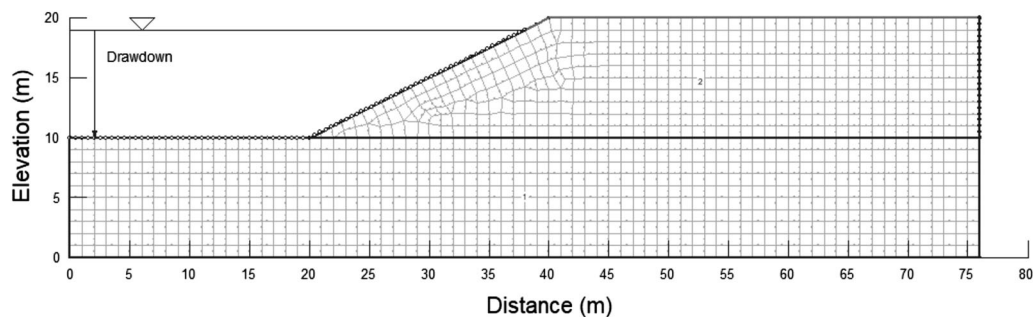


Fig. 1 Geometry of the slope and mesh of the seepage model

$$\theta = \theta_r + \frac{\theta_s - \theta_r}{\left[1 + \left(\frac{\psi}{a}\right)^n\right]^m} \quad (2)$$

where θ_r and θ_s are the residual and saturated water content, respectively, and a , n and m (with $m = 1 - \frac{1}{n}$) are curve fitting parameters. The hydraulic conductivity, which also depends on the volumetric water content, was derived by van Genuchten based on Mualem's model [36]:

$$k_r = \frac{k}{k_s} = S_e^{1/2} \left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2 \quad (3)$$

where k_r is the relative hydraulic conductivity, defined as the ratio between the unsaturated and the saturated hydraulic conductivity (k and k_s , respectively). S_e is the effective degree of saturation and is equal to:

$$S_e(\psi) = \frac{\theta(\psi) - \theta_r}{\theta_s - \theta_r}. \quad (4)$$

Figure 2 shows the SWCC and the unsaturated hydraulic conductivity curves obtained with the hydraulic parameters of Table 1. These equations are solved numerically with the finite element method implemented in SEEP/W. The resulting PWP and S_e are the inputs for the slope stability analysis.

2.2 Slope stability analysis

The factor of safety is defined as the ratio between the resistive and the active forces acting on a slip surface. The most critical slip surface that can develop within the slope will result in the minimum FOS. In this study, the FOS is calculated with the method of slices from Morgenstern–Price. The shear strength of the unsaturated soil is

predicted with the extended Mohr–Coulomb failure criterion derived by Vanapalli et al. [47]:

$$\tau = c' + (\sigma - u_a) \tan \phi' + (u_a - u_w) S_e \tan \phi' \quad (5)$$

where τ is the shear strength, σ is the total normal stress, $(u_a - u_w)$ is the matric suction ψ and c' and ϕ' are the effective cohesion and the effective friction angle, respectively. This set of analyses, i.e., transient seepage and slope stability, is repeated for a number of simulations equal to the size of the sample for each random variable.

2.3 Estimation of the probability of failure

To reduce the number of simulations, the samples of the random variables are generated with an extension of the latin hypercube sampling (LHS) with the multidimensional uniformity method [8] implemented in Python [35]. Compared to the direct Monte Carlo method, which generates pseudo-random numbers, LHS divides the range of each stochastic variable into several intervals with equal probability. This results in a better distribution of the samples in the sample space. The correlation between random variables is introduced using the methodology proposed by Iman and Conover [17]. Correlation between independent stochastic variables is induced by varying the sample ordering in each marginal distribution, while keeping the representative values untouched.

A batch script written in Python selects each combination of the generated samples and replaces the values of the random variables directly in the GeoStudio files. The batch script starts each deterministic simulation described in Sects. 2.1 and 2.2 and then reads the factor of safety from the results of each simulation. This procedure follows the nonintrusive calculation proposed by Wang et al. [50]. Finally, the probability of failure is calculated as following:

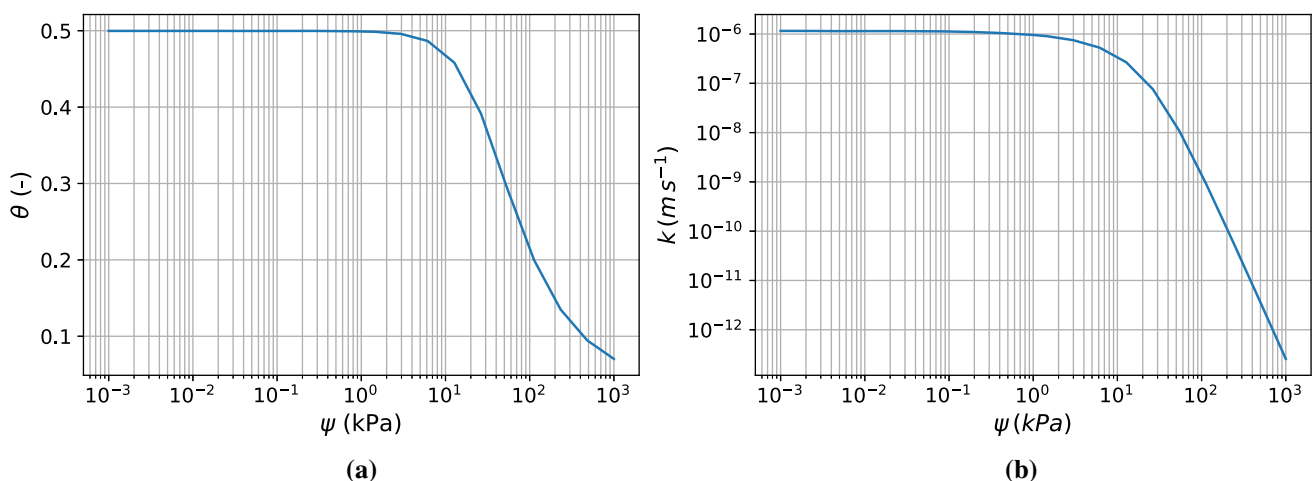


Fig. 2 Hydraulic parameters: **a** SWCC; **b** unsaturated hydraulic conductivity curve

$$p_f = \frac{1}{N} \sum_{i=1}^N I[\text{FOS} < 1] \quad (6)$$

where N is the sample size and $I[\cdot] = 1$ is an indicator function of slope failure (when failure occurs it is equal to unity).

2.4 Surrogate models

An intelligent surrogate model is adopted to enhance the computational efficiency while estimating p_f . A limited number of simulations serve as the training set, from which the model learns the underlying relationship between input parameters (e.g., shear strength and hydraulic parameters) and the output, which is the factor of safety. The problem is framed as regression, which means that a continuous value of FOS is predicted. Two predictive models from machine learning (ML) are considered in this study: support vector regression (SVR) and extreme gradient boosting (XGBoost). SVR presents several advantages. First, it has a simple implementation and a good generalization capability. Second, the model complexity of the model does not increase with the number of input variables [3]. Besides, XGBoost has high computational speed and can handle large datasets as well as missing values.

2.4.1 Support vector regression

SVR is an extension of support vector machines (SVMs) for regression problems [9]. The basic idea of SVM is that the original data are mapped into a high-dimensional feature space by a function, which can be a nonlinear transformation. A hyperplane is searched to separate the data for classification or to find the narrowest tube that contains the majority of the data for regression. The regression function can be defined as:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (7)$$

where ϕ is the nonlinear mapping function, \mathbf{w} and b are the weight vector and the offset, respectively. These parameters can be found by minimizing an objective function subject to some constraints, which aim to limit the discrepancy between the predicted and the measured output to a threshold value ϵ . A soft margin approach can be adopted to allow a number of data to be outside of the ϵ -tube by introducing slack variables. The objective function can be written as:

$$\mathcal{L}_{obj} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (8)$$

where the first term represents the complexity of the model, C is a penalty parameter and ξ and ξ^* are the slack

variables. To estimate these parameters, the objective function is minimized subject to the following constraints:

$$\begin{cases} y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i \\ \mathbf{w}^T \phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n \end{cases} \quad (9)$$

As mentioned, the function $f(x)$ can be also nonlinear and in this case, data can be mapped into a higher dimension space, called kernel space, where all the data can be linearly separated. All instances x_i are replaced with a kernel function $K(x_i, x_j)$, which is, in this study, the radial basis function (RBF):

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (10)$$

where γ is a hyperparameter that can be manually tuned. Hyperparameter tuning is then applied to search for the optimum values of the hyperparameters C and γ .

2.4.2 Extreme gradient boosting

The XGBoost algorithm is an ensemble ML algorithm based on gradient boosted decision trees. It can be used both for regression and classification problems. This model, which was developed by Chen and Guestrin [6], recently became very popular for winning several ML competitions [4] because of its high execution speed and model performance. Ensemble decision trees are the base learners that predict the output (which is in this study FOS) by means of K additive functions:

$$\hat{y}_i = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F} \quad (11)$$

where f_k is the function describing one independent tree, \mathcal{F} is the space of all possible regression trees, \mathbf{x}_i are the input features and \hat{y}_i are the predicted output. In supervised learning, the training phase serves to fit a model that relates the output to the input features with the best fit. The goodness of fitting can be measured with an objective function that consists of two terms: the training loss (e.g., the mean squared error) and the regularization term, which controls model complexity (hinders overfitting), as described in Eq. (12) [6].

$$\mathcal{L}_{obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (12)$$

The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (13)$$

where w represents the score on a leaf and T is the number

of leaves in the tree [6]. An additive strategy is applied to train the model; each new tree learns from the previous one and the prediction value at the time step t would be:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(\mathbf{x}_i) = \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i) \quad (14)$$

The objective function at the time step t is defined as:

$$\mathcal{L}_{obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t) \quad (15)$$

The loss function can be approximated to the second-order Taylor series expansion and the objective function can be rewritten:

$$\mathcal{L}_{obj}^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t) \quad (16)$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ are the partial derivatives of the first and second order of the loss function. The final form of the optimization function can be obtained by removing all the constant terms:

$$\mathcal{L}_{obj}^{(t)} = \sum_{i=1}^n [g f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \Omega(f_t) \quad (17)$$

2.4.3 Preprocessing

The available dataset is divided into training and test set, with a ratio of 70–30%, which is a compromise between the subdivision in two thirds–one third [27] and the Pareto principle (also known as 80–20 rule). This ratio is a common choice in machine learning practice, although other train–test ratios are also available (e.g., 60–40%, 80–20% and 90–10%) [46]. However, the optimal train–test split ratio can be generally determined with a trial-and-error procedure [37]. In this study, the performance of the surrogate model with the selected train–test ratio has been compared with the following ones: 60–40%, 80–20% and 90–10%. For all tested split ratios, the performance in both training and test shows negligible discrepancy, meaning that the split ratio has a little influence on the predictive performance of the surrogate model. Cross-validation (CV) is applied to prevent overfitting: the training set is split into k groups or folds and the model is trained on $k - 1$ folds and validated on the remaining one. After cross-validation, a mean model performance score over all folds is provided. This process can be combined with the model selection phase, when the hyperparameters that give the best performance in terms of a prediction score are selected. For the SVR, the optimization of the hyperparameters is

performed with two techniques, namely the randomized search and the Bayes search, which are part of the scikit-learn [39] and the scikit-optimize libraries [14], respectively. In the randomized search, a set of values is randomly sampled from a statistical distribution for each hyperparameter, i.e., C and γ . Since these hyperparameters take continuous values, a log-uniform distribution is chosen to generate random samples of the parameter space. Finally, the algorithm searches the best values for the hyperparameters. Similarly, the Bayes search samples possible hyperparameters values from a given distribution. Based on Bayes' theorem, this algorithm uses the acquired information to find the values of the hyperparameters that are likely to increase the model's performance.

2.5 Stochastic parameters

In this study, the effective cohesion (c'), the effective friction angle (φ'), the saturated hydraulic conductivity (k_s) and the curve fitting parameters of the SWCC (a and n) are modeled as random variables. The mean values (μ), the coefficients of variation (COV) and correlation (ρ) are shown in Table 1. The random variables have a lognormal distribution to avoid unrealistic negative values. A lower bound of 1.05 is imposed to the parameter n to fulfill the condition $0 \leq S_e \leq 1$ [22] and an upper bound of 45° is imposed to the parameter φ' to discard values of the effective friction angle that deviate from the geotechnical practice. The samples from a truncated lognormal distribution are generated with the R package EnvStats [33]. A negative correlation is imposed between the effective cohesion and the effective friction angle as well as between the parameters a and n of the SWCC. Phoon et al. [41] found out that among several classical distributions (e.g., normal, gamma, exponential) the lognormal distribution is the most suitable to describe the variability of the curve fitting parameters a and n of the SWCC. Typical values of the COV of the shear strength and hydraulic parameters are found in [22, 40]. Phoon et al. [41] investigated the correlation between a and n for different types of soil and found out that these two parameters are negatively correlated with each other, with a range of $\rho_{a,n}$ from -0.5 to -0.25% .

The sample size of the shear strength and hydraulic parameters used for training the surrogate model is 2000 samples. Running 2000 simulations with GeoStudio on a desktop computer with a processor speed of 2.40 GHz and a RAM memory of 8GB took approximately 18 hours. Figure 16 in Appendix shows the distribution of the samples and the correlation between the parameters c' and φ' and between a and n . An overview of the implementation procedure is depicted in Fig. 3.

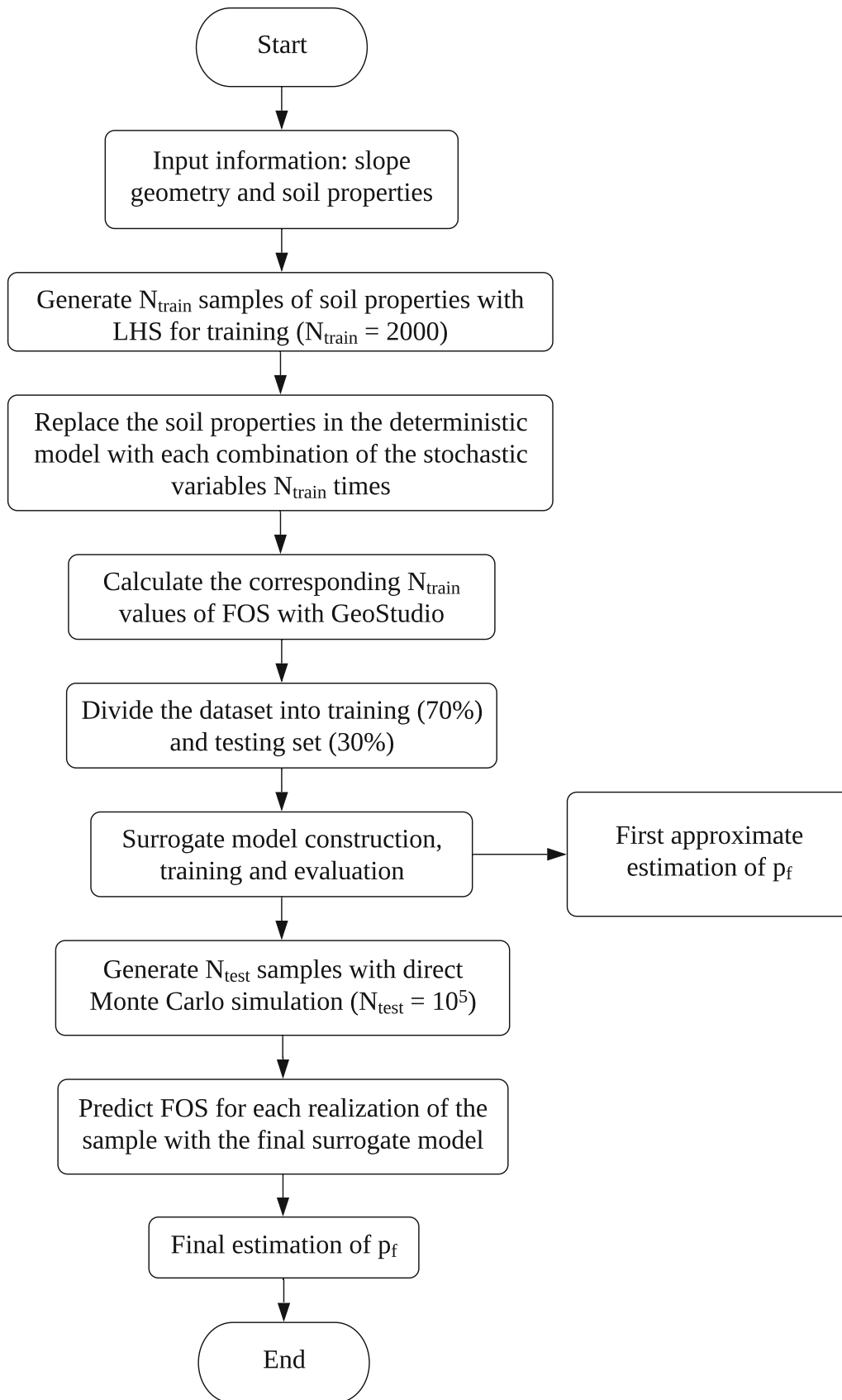


Fig. 3 Implementation procedure for the estimation of p_f with surrogate model

3 Results

The seepage analysis and the slope stability analysis are conducted for each sample combination (i.e., 2000 times) and for each time step of the transient seepage analysis. The drawdown rate is 1.8 m d^{-1} : the reservoir level reaches 10 m elevation after 5 days. In Fig. 4, the distribution of the PWP is shown before and after drawdown. The PWP and the effective degree of saturation are the inputs of the slope stability analysis and the FOS is calculated with the Morgenstern–Price method. Table 2 illustrates the FOS calculated with the LEM at each time step. These results are obtained by setting the mean values of the shear strength and hydraulic parameters of Table 1.

A first estimation of p_f can be achieved according to Eq. (6). Figure 17 in Appendix shows the probability of failure with increasing number of simulations at each time step. However, the number of samples provided may not be sufficient to accurately estimate p_f , especially for very low values of p_f . The surrogate model is tested on a larger sample to overcome this limitation. A direct Monte Carlo simulation generates 10^5 combinations of input parameters. Instead of calculating the FOS with GeoStudio for each combination, the surrogate model is employed to predict it. The input parameters (also called “features” in machine learning) are the shear strength and the hydraulic parameters. FOS is the output parameter, which is calculated at each time step: this means that the surrogate model is re-trained at each time step of the simulation. Finally, a more accurate estimation of p_f is obtained from Eq. (6).

Two methods from ML are tested, namely SVR and XGBoost. For the former algorithm, two different optimization techniques for hyperparameter tuning are applied and compared: randomized search and Bayes search as described in Sect. 2.4. A fivefold cross-validation is applied both to SVR and XGBoost. Table 3 shows the error metrics for the test set in terms of coefficient of determination (R^2), Root mean squared error (RMSE) and Mean Absolute Error (MAE). SVR reaches high accuracy with both optimization techniques. However, the XGBoost method has superior performance, with a test R^2 higher than 98.7%, and it is also computationally more efficient. Therefore, the XGBoost algorithm is selected for the estimation of p_f and for a sensitivity analysis of the stochastic parameters. Figure 5 compares the FOS calculated with the Morgenstern–Price method, and the values obtained with the XGBoost model for the test set (i.e., 30% of the 2000 samples) of the time step no. 0. The predicted values are in

good agreement with those calculated by the GeoStudio software, showing that the XGBoost model is capable of predicting the factor of safety with high accuracy.

The XGBoost model is used to predict FOS in 4 different scenarios of drawdown rate v : 0.5 m d^{-1} , 1 m d^{-1} , 1.8 m d^{-1} (as set in Sect. 3) and 3 m d^{-1} . The water level of the reservoir linearly decreases with time, until it reaches 10 m elevation after 18 days in scenario 1, after 9 days in scenario 2, after 5 days in scenario 3 and after 3 days in scenario 4 (Fig. 6). For each scenario, the XGBoost model is trained on a different dataset: the input parameters (c' , ϕ' , k_s , a , n) are sampled according to the statistics in Table 1. The output of the model (i.e., FOS) is calculated at each time step with GeoStudio. It takes approximately 30 s to train the XGBoost model for each scenario. Figure 6 shows FOS obtained with the mean values of the soil parameters in Table 1 and p_f , which is estimated with the predictions of the XGBoost model (through Eq. 6).

As shown in Fig. 18 in Appendix, p_f converges to a stable value after approximately 60,000 simulations except the time steps no. 0, 1 and 2, where p_f is less or approximately equal to 1%. In these last two cases, p_f does not converge after 100,000 simulations, indicating that a higher number of simulations would be necessary for a more accurate estimation. The uncertainty related to the estimated p_f can be obtained as follows: $\text{COV}_{p_f} = \sqrt{\frac{1-p_f}{n p_f}}$ [2], where $n = 10^5$. For time step 1 (after 12 h from rapid drawdown), this value is equal to 31.62%, which is slightly above the target COV_{p_f} of 30% [19, 40] indicated in the literature. Therefore, the potential inaccuracy of the results at time step 1 should be treated with caution. At time step 2 (after 24 h), COV_{p_f} is equal to 2.46%, much lower than the suggested threshold and, consequently, the estimated value of p_f can be considered acceptable.

4 Discussion

A minimum factor of safety of 1.5 should be satisfied in steady-state conditions [16], but a lower FOS of 1.1–1.2 can be assumed acceptable under rapid drawdown [28, 45]. As expected, a higher drawdown velocity implicates a lower critical FOS and the variation of p_f with time mirrors this behavior. Siacara et al. [45] verified that the critical time in terms of minimum FOS and maximum p_f do not necessarily coincide. In this study, this effect may not be

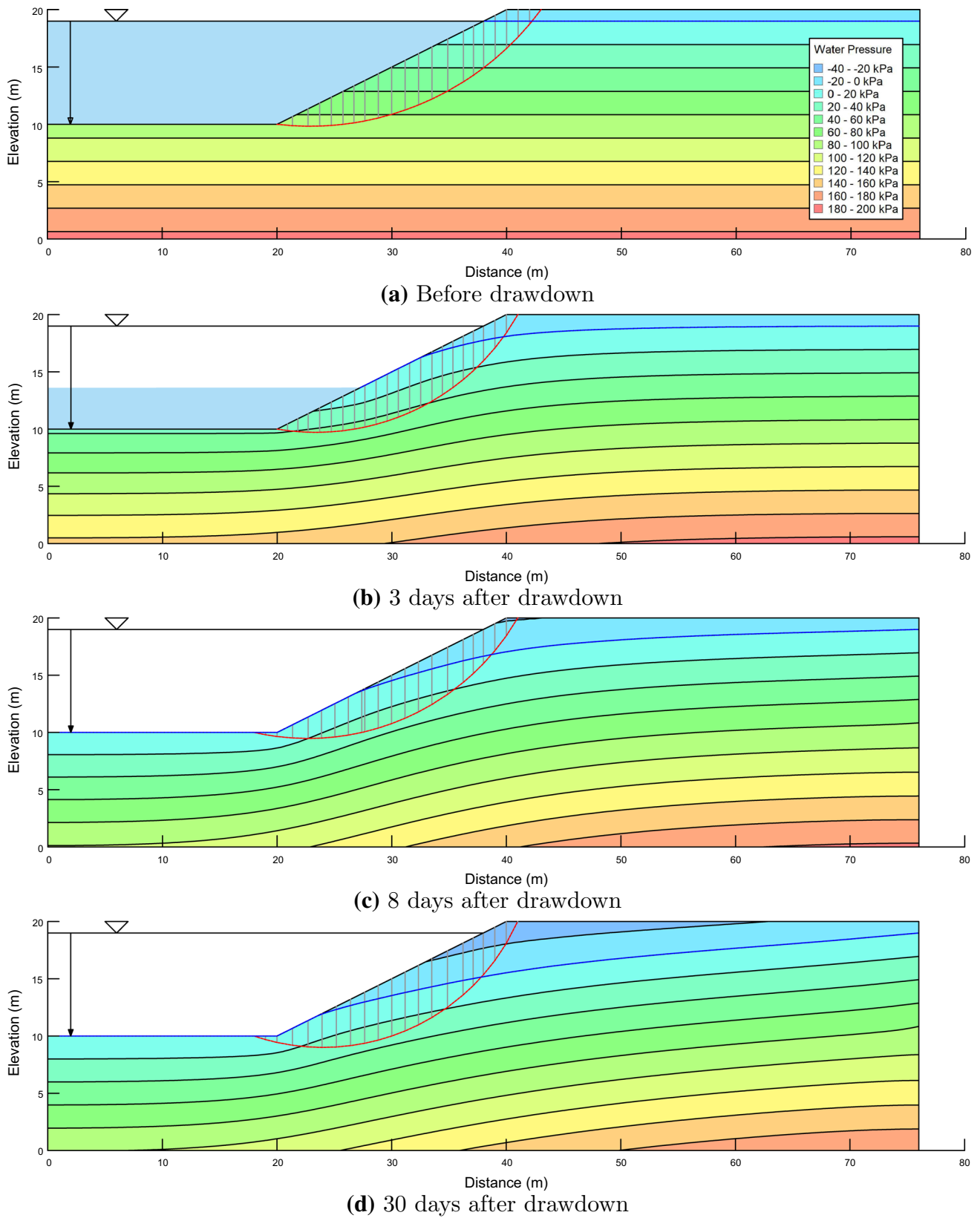


Fig. 4 Distribution of the PWP in the slope (the dashed red and blue lines represent the critical slip surface and the water table, respectively)

Table 2 FOS for each time step calculated with the mean values of the parameters

Time step no.	Time (days)	FOS
0	0	1.81
1	0.5	1.55
2	1	1.38
3	2	1.16
4	3	1.03
5	5	1.02
6	8	1.13
7	11	1.18
8	15	1.23
9	22	1.27
10	30	1.31

visible due to the frequency of the calculation steps in the seepage and slope stability analysis. However, according to the mentioned study, the critical FOS arises before the critical p_f , with 24 hours of delay for a drawdown rate of 2 m d^{-1} and 3 days for 0.5 m d^{-1} . Following this reasoning, the peak in p_f (Fig. 6) may occur slightly later in time, which could be verified by increasing the frequency of the calculations steps. As shown in Fig. 6, the peak of p_f occurs when the reservoir level is stable at 10 m elevation or at the previous time step. In the first ($v = 0.5 \text{ m d}^{-1}$) and second scenario ($v = 1 \text{ m d}^{-1}$), this corresponds to 15 days with $p_f = 22.81\%$ and to 8 days with $p_f = 38.04\%$, respectively. In the third scenario (whose results are also shown in Figs. 17 and 18), the peak of p_f takes place exactly when the minimum water level is

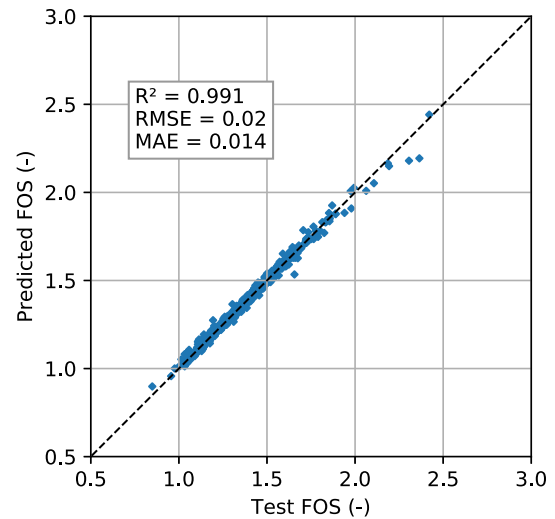


Fig. 5 Validation of the XGBoost model on the test set of time step no. 2. The dashed line represents a 45° line

reached, that is, after 5 days with $p_f = 48.72\%$. The last scenario has the highest drawdown rate ($v = 3 \text{ m d}^{-1}$): the maximum p_f occurs after 2 days with $p_f = 72.42\%$, while the water level is constant at 10 m starting from 3 days after drawdown.

The feature importance is calculated for the test set using XGBoost. This parameter is defined as the variation in a model performance score (in this study R^2) after random permutation of a single feature. If the model performance score drops significantly after the random shuffle of a particular feature, the model is thus greatly dependent on this one. Figure 7 shows the feature importance of the effective friction angle ϕ' , the effective cohesion c' and the curve fitting parameters a and n . The shear strength

Table 3 Error metrics of the surrogate models for the test set

Time step no.	SVR			SVR			XGBoost		
	(Randomized search)			(Bayes search)					
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
0	0.982	0.040	0.031	0.979	0.044	0.036	0.993	0.024	0.015
1	0.978	0.037	0.029	0.978	0.037	0.030	0.992	0.022	0.015
2	0.975	0.035	0.026	0.976	0.034	0.026	0.991	0.020	0.014
3	0.952	0.040	0.031	0.958	0.037	0.030	0.990	0.018	0.012
4	0.941	0.039	0.031	0.943	0.038	0.031	0.987	0.018	0.012
5	0.960	0.034	0.026	0.962	0.033	0.026	0.992	0.015	0.010
6	0.970	0.035	0.027	0.966	0.037	0.031	0.990	0.020	0.013
7	0.976	0.033	0.026	0.964	0.041	0.034	0.991	0.021	0.014
8	0.980	0.033	0.025	0.979	0.033	0.025	0.989	0.024	0.016
9	0.981	0.033	0.025	0.982	0.033	0.024	0.988	0.026	0.017
10	0.980	0.035	0.026	0.979	0.036	0.030	0.990	0.026	0.017

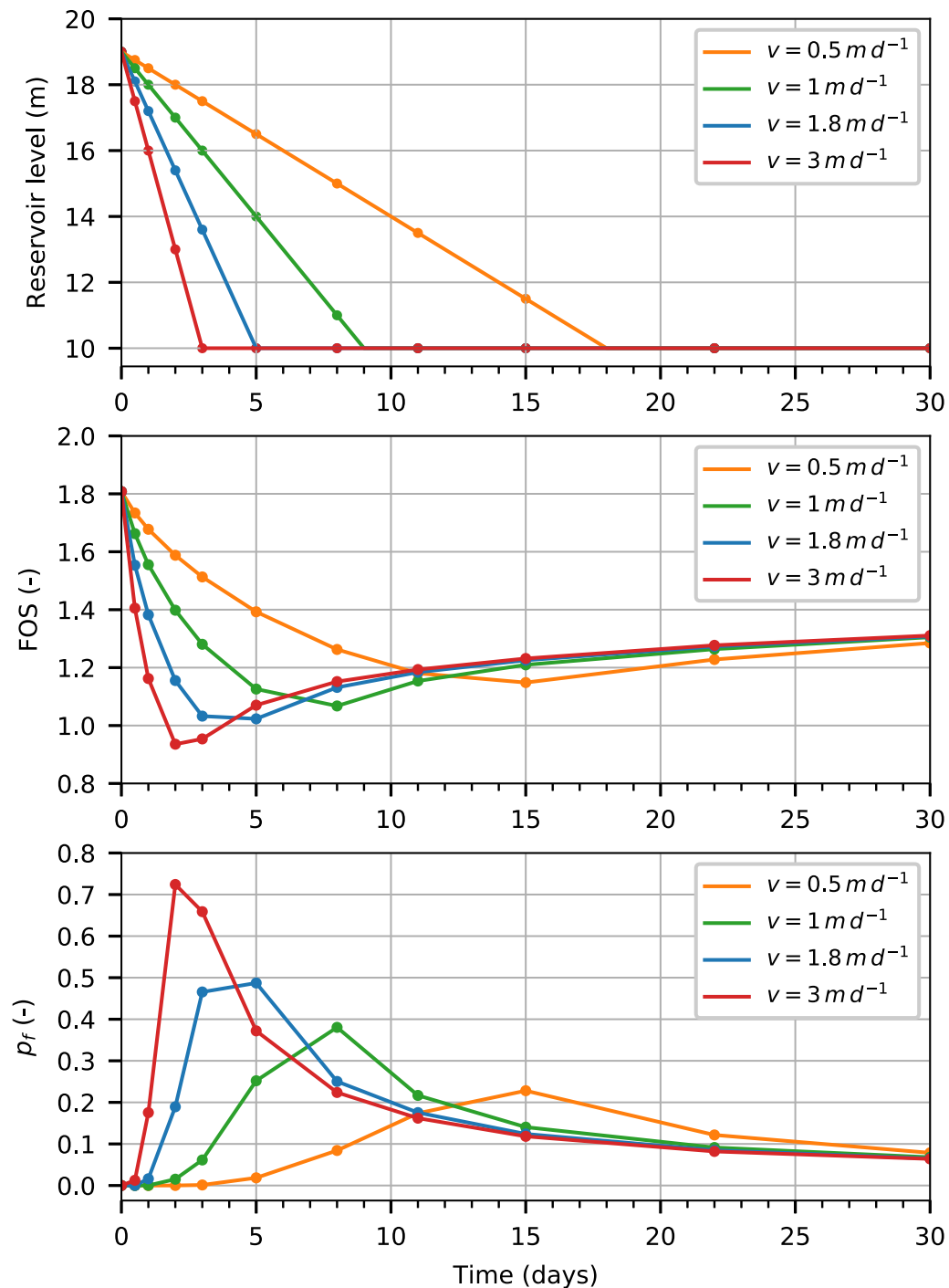


Fig. 6 Variation of the reservoir water level, FOS and p_f with time for different drawdown rates

parameters exhibit the highest feature importance, with values that range from 76 to 92% for φ' and from 5 to 23% for c' . An interesting aspect that emerges from Fig. 7 is the variation in the feature importance with time. For each time step, the surrogate model is trained with the same features (i.e., the samples of the random variables) but with different data for the label (i.e., FOS). Thus, the feature

importance will change with each time step and consequently with the reservoir level. In general, the feature importance of φ' decreases for the duration of the drawdown, and then, it increases as the reservoir level is constant. A visible drop occurs earlier in time with increasing drawdown rate. The reversed trend is observed for c' . As the water level decreases, the feature importance of c'

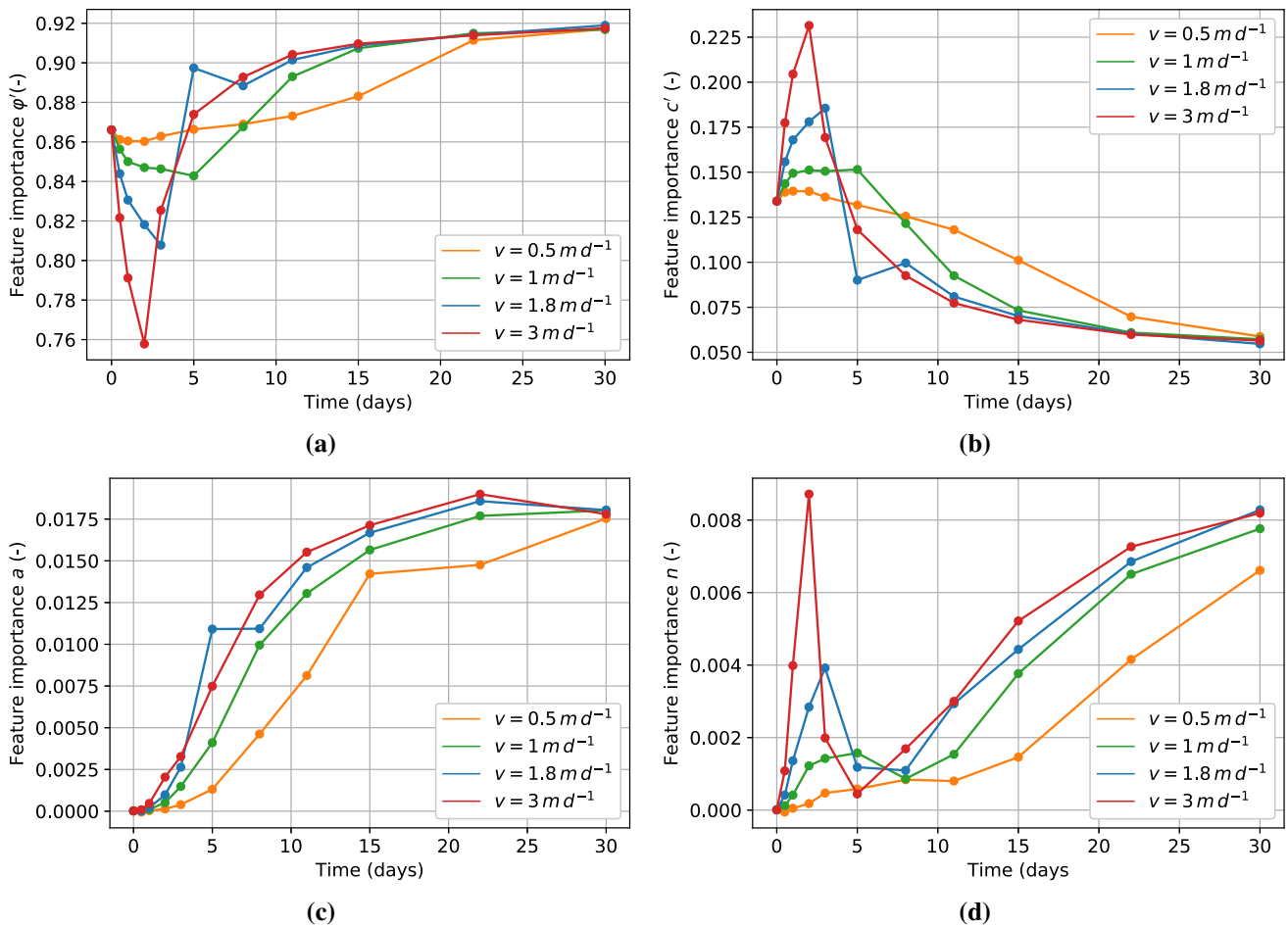


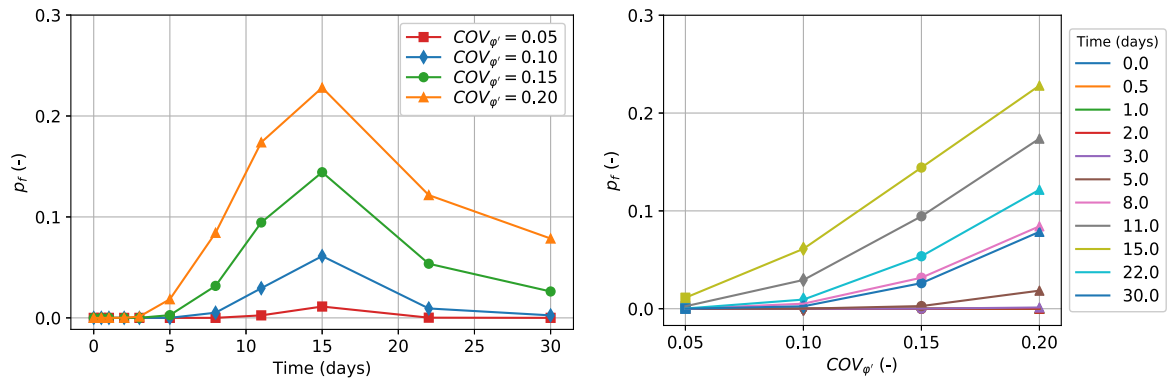
Fig. 7 Feature importance of the input parameters: **a** ϕ' ; **b** c' , **c** a ; **d** n

increases until a peak is reached. Similar to the previous case, the peak is found earlier in time as the drawdown rate increases. The feature importance of the curve fitting parameters of the SWCC is much lower, below 2% for a and below 0.1% for n . A general increase in the feature importance of a and n can be observed after rapid drawdown, when the reservoir level is stable at 10 m. The parameter a increases asymptotically with time, while n has a peak approximately at the end of the drawdown, whose magnitude increases with higher drawdown rate. The feature importance of k_s is omitted because lower than 0.5%. The observed results are interpreted as follows. With reference to Eq. (5), the shear strength can be divided into three components: cohesive, frictional and suction strength. As the reservoir level drops, the water load on the slope surface decreases together with the frictional strength along the slip surface. When the reservoir level becomes stable at 10 m, the pore water pressure in the slope dissipates and the frictional strength rises, reflecting the variation in the feature importance of the ϕ' . Regarding the curve fitting parameters a and n , the suction strength slowly increases with time, thus resembling the pattern of their feature

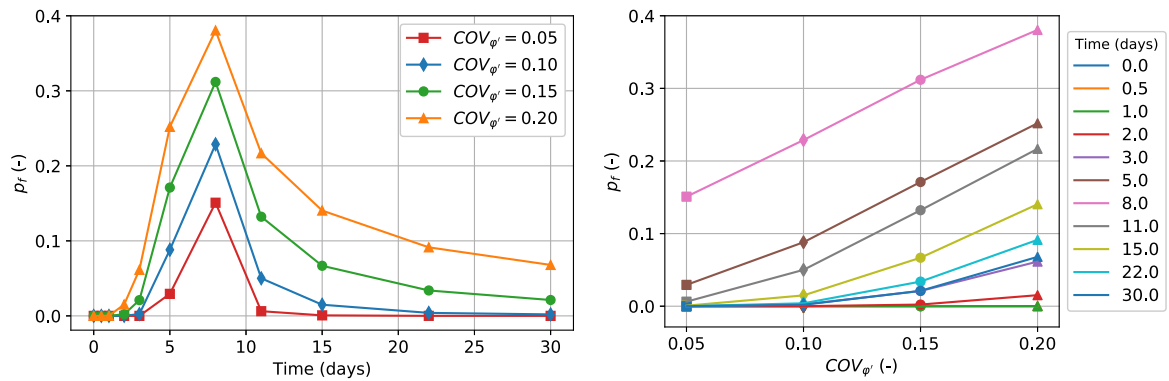
importance. However, it is important to point out that the correlation between the random variables can influence the results of the feature importance analysis. When the instances of a feature are permuted, the relationship between features and label (or input and output parameters) can be still obtained from the correlated feature [39]. This could explain the variation in the feature importance of c' that mirrors the one of ϕ' , although the cohesive strength is

Table 4 Range of the coefficients of variation and correlation for the sensitivity analysis

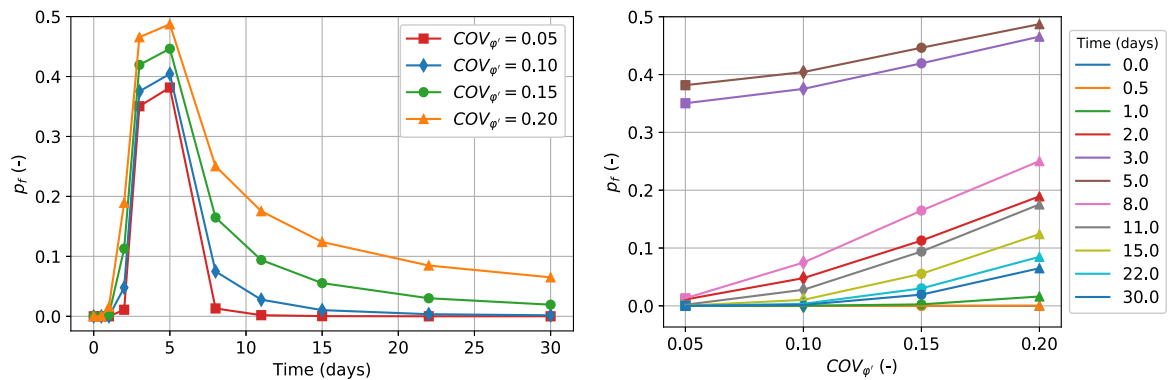
Parameter	Range
$COV_{c'}$	0.2–0.5
$COV_{\phi'}$	0.05–0.2
COV_{k_s}	0.3–0.9
COV_a	0.2–0.8
COV_n	0.1–0.4
$\rho_{c',\phi'}$	– 0.5–0.5
$\rho_{a,n}$	– 0.5–0.5



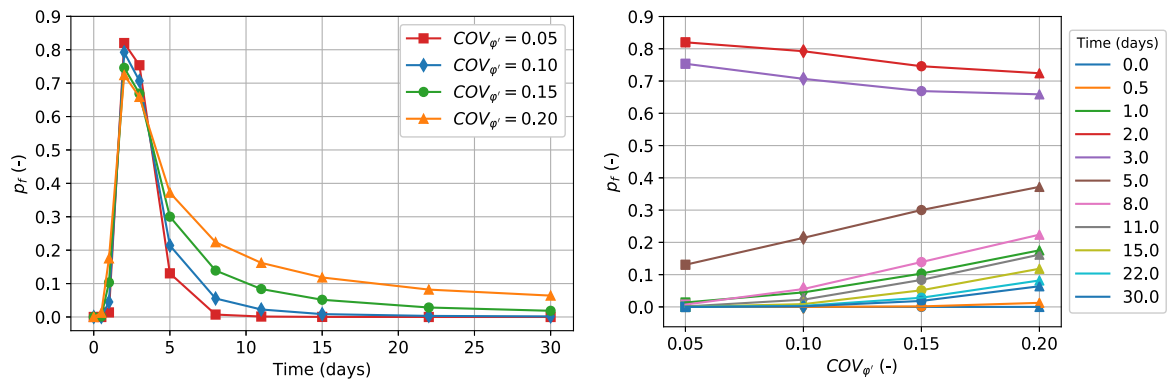
(a) Scenario 1: $v = 0.5 \text{ m d}^{-1}$



(b) Scenario 2: $v = 1 \text{ m d}^{-1}$



(c) Scenario 3: $v = 1.8 \text{ m d}^{-1}$



(d) Scenario 4: $v = 3 \text{ m d}^{-1}$

Fig. 8 Variation of p_f with $COV_{\phi'}$

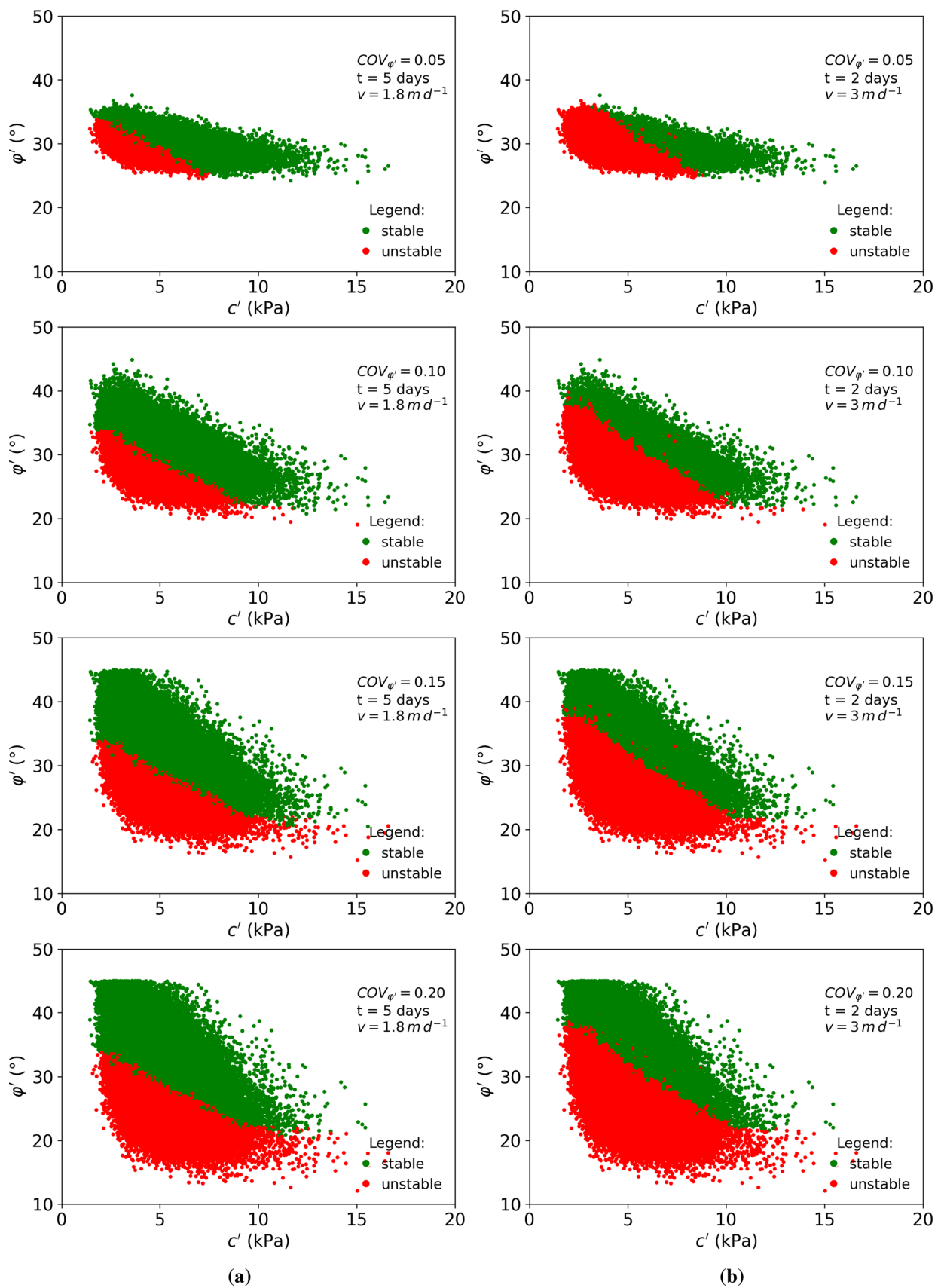


Fig. 9 Variation of p_t with $COV_{\phi'}$: **a** scenario 3 ($v = 1.8 \text{ m d}^{-1}$) after 5 days; **b** scenario 4 ($v = 3 \text{ m d}^{-1}$) after 2 days

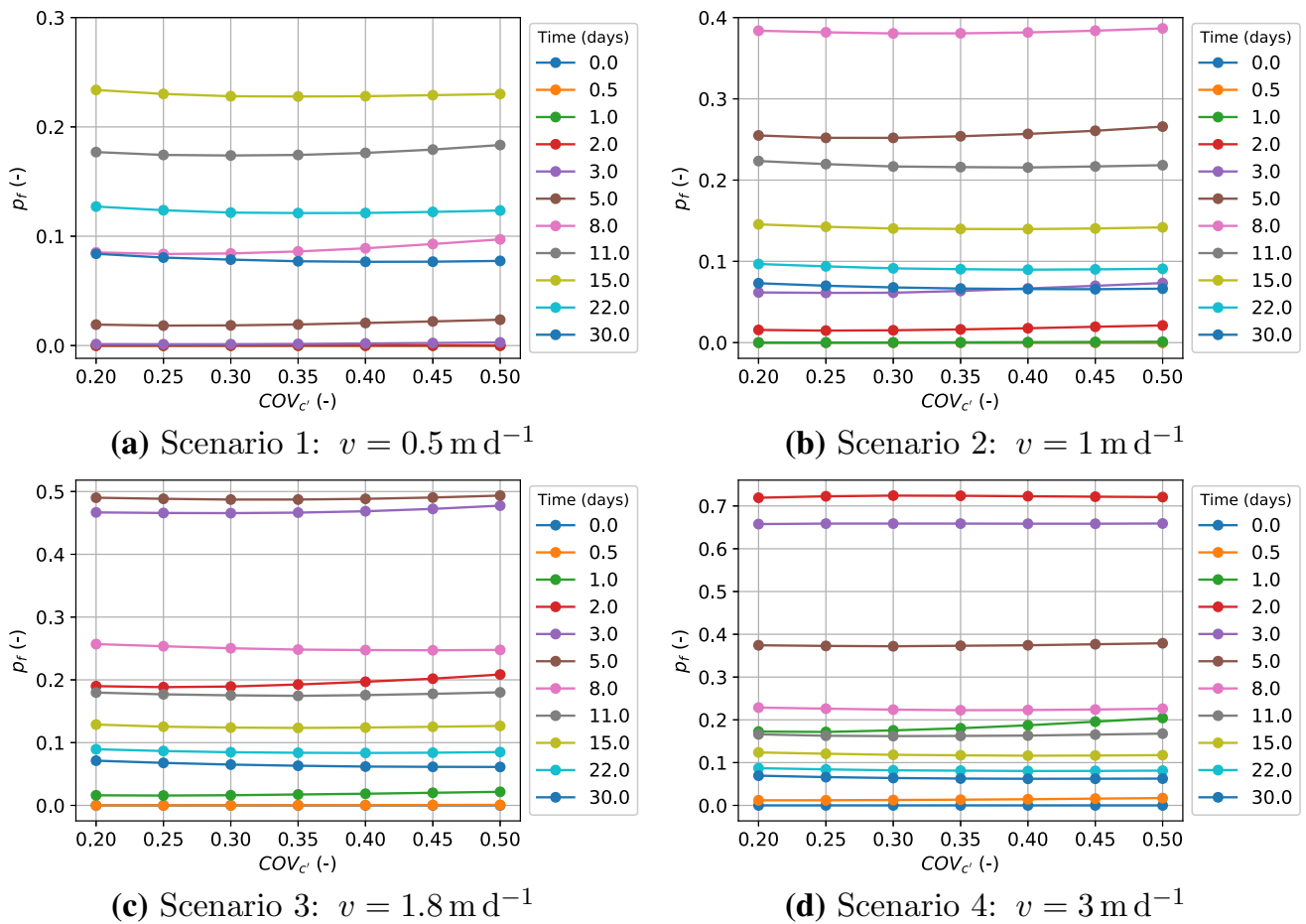


Fig. 10 Variation of p_f with $COV_{c'}$

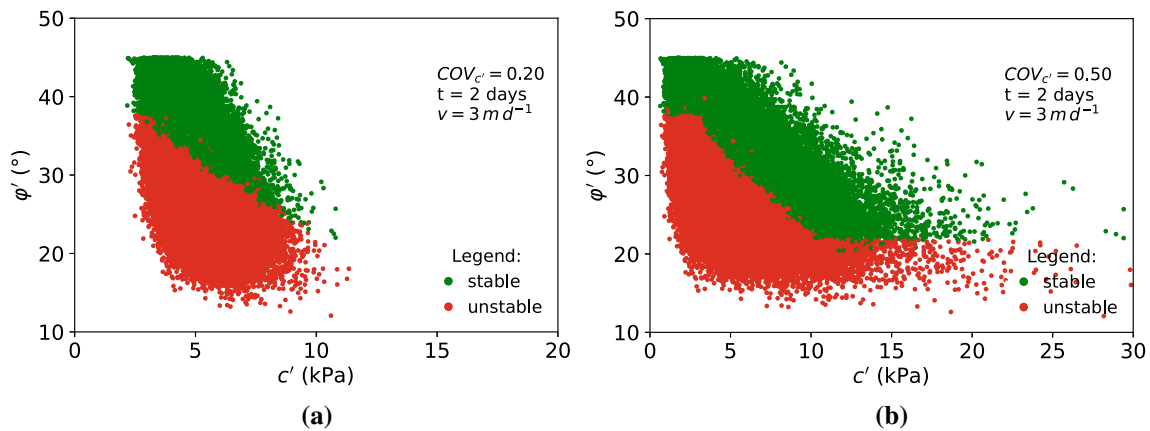
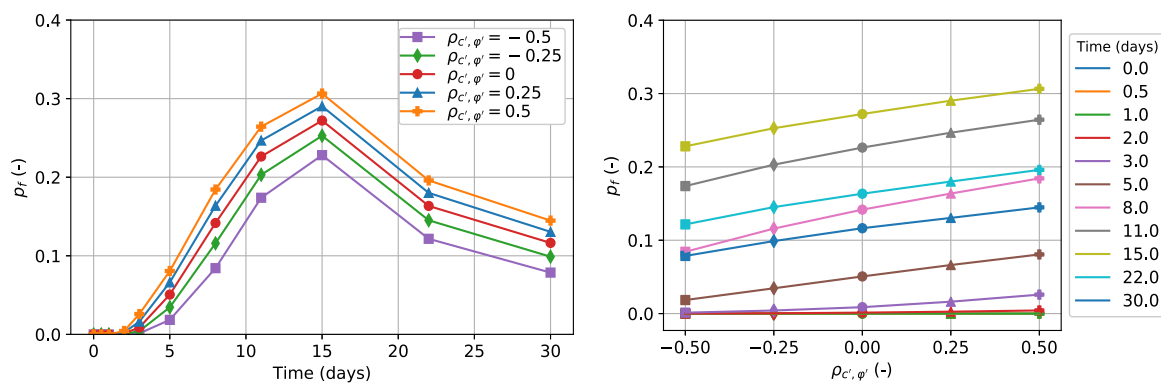


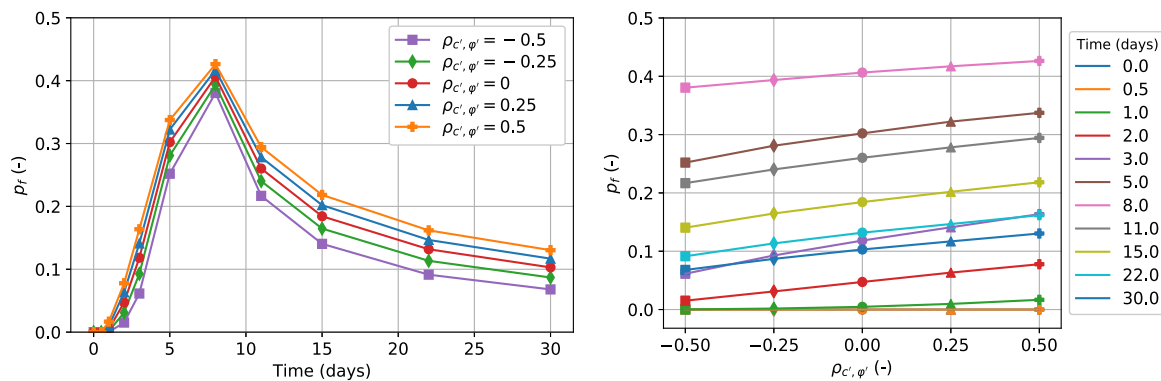
Fig. 11 Variation of p_f with $COV_{c'}$: **a** $COV_{c'} = 0.20$ and **b** $COV_{c'} = 0.50$

constant with time. Similar outcomes in terms of relative importance between features are presented by Wang et al. [51]. The authors evaluate the feature importance of the shear strength and hydraulic parameters of a surrogate model based on MARS. This model is applied in the probabilistic analysis of an earth dam slope under transient

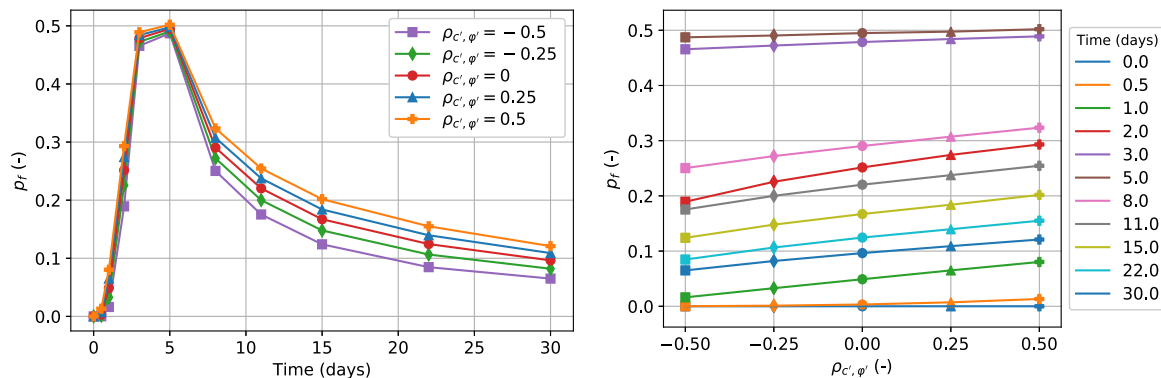
seepage. Although the absolute values of the feature importance are not directly comparable with those of this study, a general trend in terms of relative importance of the individual features can be identified. The results of Wang et al. [51] show that the shear strength parameters have the



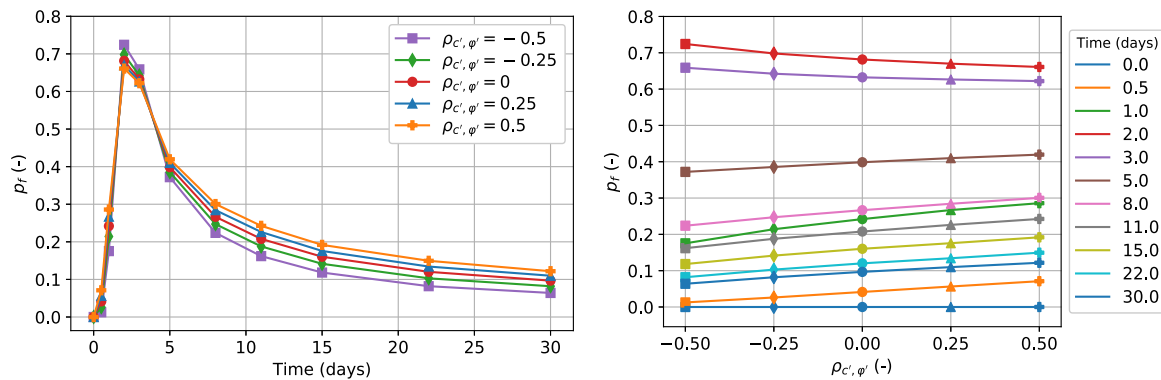
(a) Scenario 1: $v = 0.5 \text{ m d}^{-1}$



(b) Scenario 2: $v = 1 \text{ m d}^{-1}$



(c) Scenario 3: $v = 1.8 \text{ m d}^{-1}$



(d) Scenario 3: $v = 3 \text{ m d}^{-1}$

Fig. 12 Variation of p_f with $\rho_{c',\varphi}$

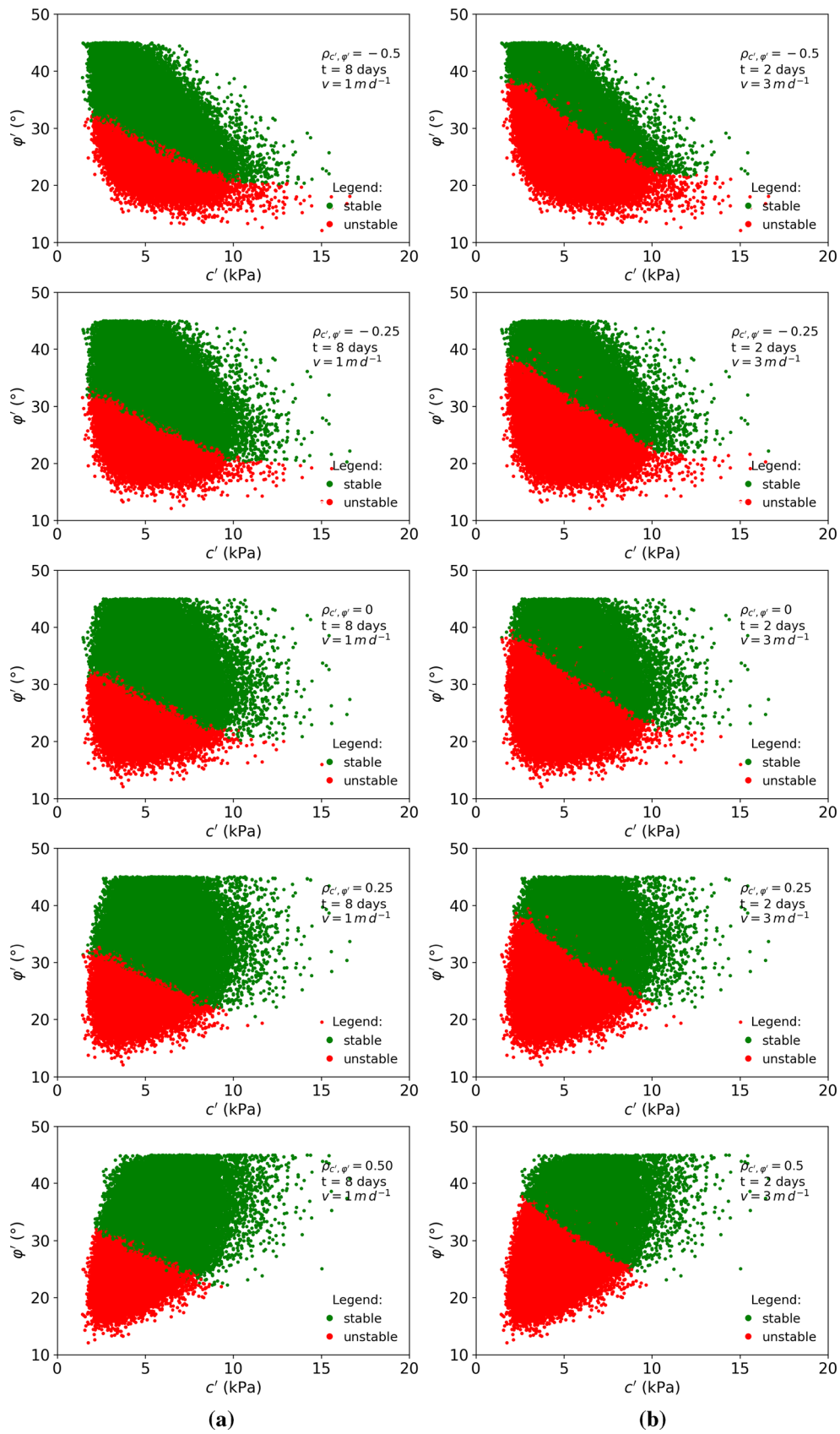
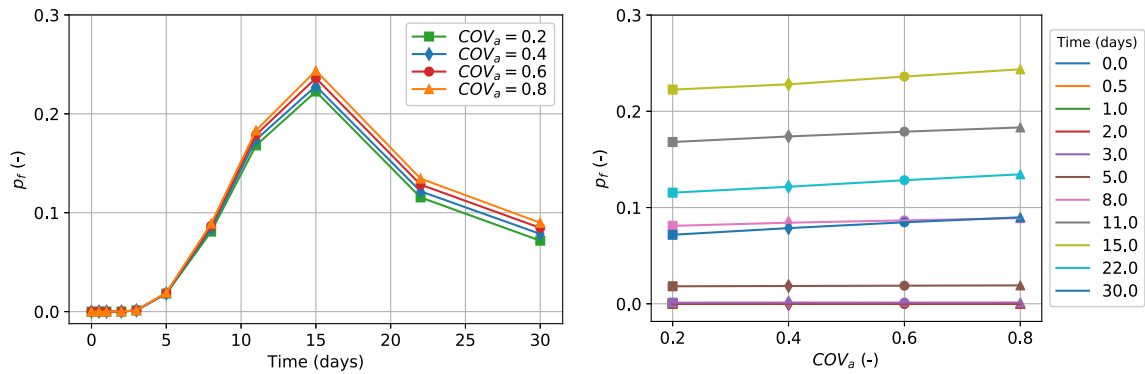
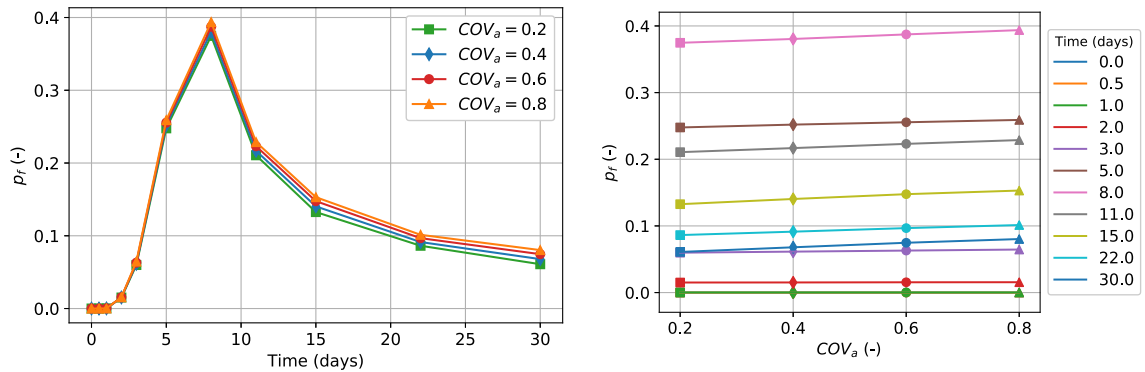


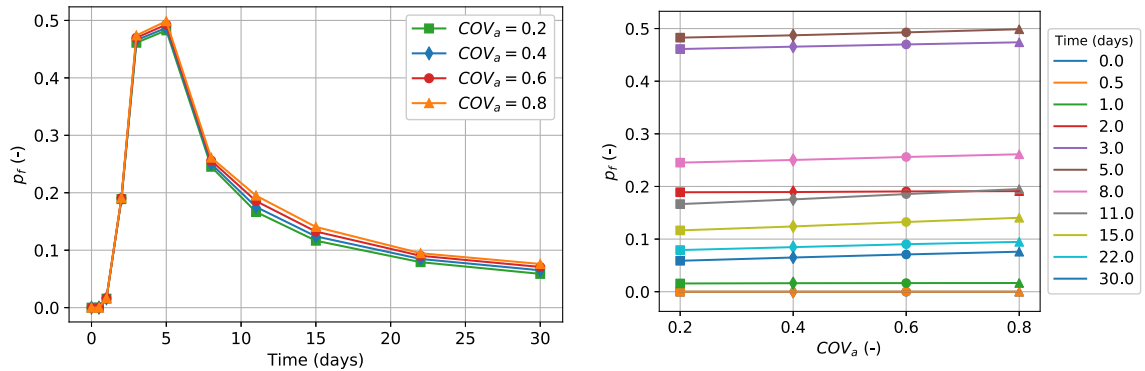
Fig. 13 Variation of p_T with $\rho_{c', \phi'}$: **a** scenario 2 ($v = 1 \text{ m d}^{-1}$) after 8 days; **b** scenario 4 ($v = 3 \text{ m d}^{-1}$) after 2 days



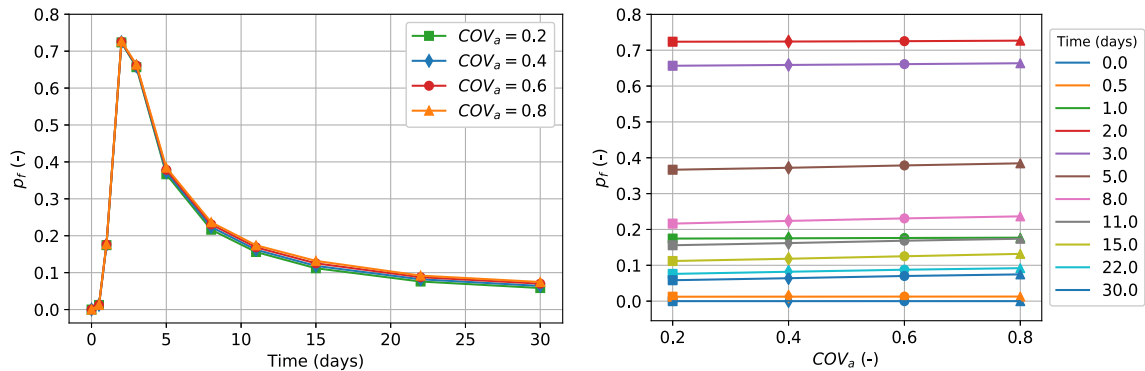
(a) Scenario 1: $v = 0.5 \text{ m d}^{-1}$



(b) Scenario 2: $v = 1 \text{ m d}^{-1}$



(c) Scenario 3: $v = 1.8 \text{ m d}^{-1}$



(d) Scenario 4: $v = 3 \text{ m d}^{-1}$

Fig. 14 Variation of p_f with time and COV_a

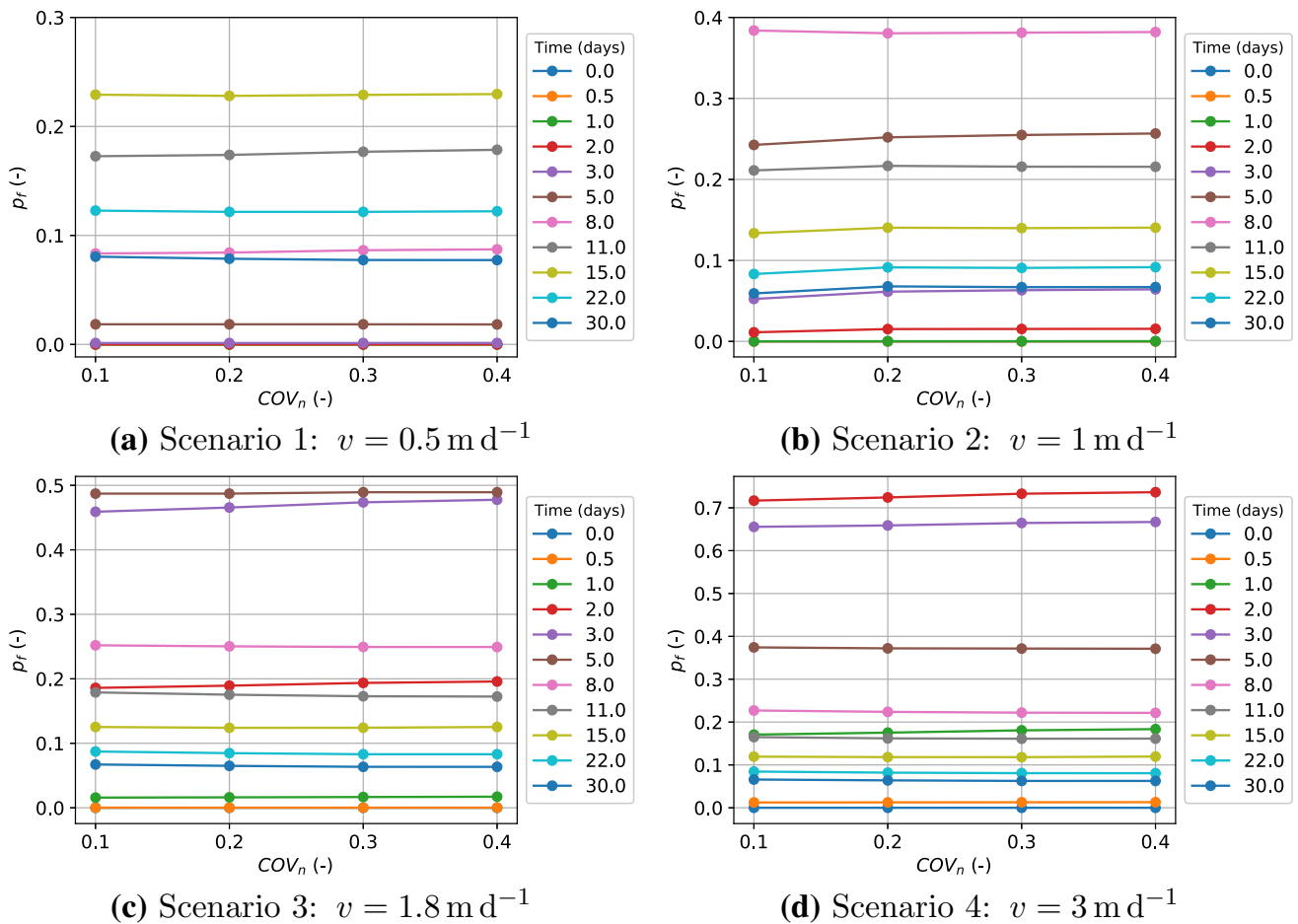


Fig. 15 Variation of p_f with COV_n

highest importance, which is approximately ten times higher than that of the hydraulic parameters (a , n and k_s).

4.1 Sensitivity analysis

A sensitivity analysis is conducted to study the effect of the uncertainty of each soil parameter regarded in this study as stochastic (i.d. c' , ϕ' , k_s , a and n) and their correlation ($\rho_{c',\phi'}$ and $\rho_{a,n}$) on p_f . The considered ranges of COV (Table 4) are similar to that reported by Jiang et al. [22]. As the COV of one parameter or the coefficient of correlation is modified, the other parameters are untouched and take the same values as shown in Table 1. A new dataset is generated with a direct Monte Carlo simulation (10^5 samples) according to the variation in the COV or of the coefficient of correlation that is of interest. For each scenario and each time step of the simulation, an XGBoost model previously trained (Sect. 3) is saved and tested on the new sample. This operation takes approximately 65 s, with a total time of 1 h required to estimate p_f for all the cases included in the sensitivity analysis.

4.2 Effect of COV of the shear strength parameters on the probability of failure

The effect of the uncertainty of c' and ϕ' is investigated for all scenarios. Figure 8 shows the variation of p_f with time (left side) and with increasing $COV_{\phi'}$ (right side) for the different scenarios presented above. The general trend is of increasing p_f with increasing $COV_{\phi'}$, which was also found by Jiang et al. [22], Wang et al. [51]. The maximum gain in p_f is found in the scenarios 3 and 4 (Fig. 8c, d) at the time steps 8 and 5 days, respectively, with an increase of 23.72% and 24.14% due to $COV_{\phi'}$ varying from 0.05 to 0.20. In scenario 2, the highest increment of p_f (Fig. 8b) is slightly above 22% at time steps 5 and 8 days.

An interesting aspect to look at is that, when the absolute value of p_f is at its peak (time steps 15, 8, 5 and 2 days in scenarios 1, 2, 3 and 4), the increase of p_f reduces between scenarios with increasing drawdown rate. The extreme case is observed in scenario 4 (Fig. 8d), where p_f reduces with increasing $COV_{\phi'}$ at time steps 2 and 3 days. The reason for this is explained in Fig. 9, where the safety and failure regions are shown for scenario 3 after 5 days

(on the left side) and scenario 4 after 2 days (on the right side). These conditions represent the peaks of p_f in these two scenarios. The red and the green points represent slope failure and safety, respectively. In scenario 3, p_f is lower than 50% and the mean value of FOS, which is the center of gravity of the data, is in the safety region (left side of Fig. 9). When $COV_{\varphi'}$ increases, the datapoints spread in the direction of φ' and more datapoints belong to the failure region. In scenario 4 (right side of Fig. 9), the mean of FOS is lower than 1 and the center of gravity of the data lies in the failure region. In this case, the reversed behavior is observed: as $COV_{\varphi'}$ increases more datapoints will belong to the safety region. This aspect was also observed by Griffiths and Fenton [12], Javankhoshdel and Bathurst [18] showing that for slopes with FOS lower than 1, p_f decreases with increasing COV. Generally speaking, a low COV or a low standard deviation imply that the data are more closely distributed to their mean values. When COV increases, the likelihood that a realization belongs to the opposite region of that of the mean value is higher.

The effect of $COV_{c'}$ on p_f is less significant than of the effective friction angle. This is probably due to the type of soil that has been considered in this study (e.g., clayey sand), where the contribution of the cohesion to the shear strength is limited. The variation of p_f with $COV_{c'}$ is very small for all scenarios, always lower than 4%. In Fig. 10, an increasing trend can be identified predominantly at the time steps that correspond to a reduction in the water level. In scenario 1, this corresponds to the time steps from 0 to 11, in scenario 2 from 0 to 5 days, in scenario 3 from 0 to 3 days and in scenario 4 from 0 to 1 day. A positive correlation between p_f and $COV_{c'}$ was also found by Jiang et al. [22]. However, when the water level is constant at 10 m, p_f slightly decreases. As already mentioned, the effect of $COV_{c'}$ on p_f is very limited if compared with that of $COV_{\varphi'}$. Figure 11 shows that the increase of $COV_{c'}$ produces an expansion of both safety and failure regions in the direction of the effective cohesion. For a $COV_{c'} = 0.50$, the limit between the safety and failure region becomes a horizontal line after 10 kPa. This means that for an effective cohesion higher than 10 kPa the influence of c' on p_f is very small or negligible.

4.3 Effect of the correlation of the shear parameters

The influence of $\rho_{c',\varphi'}$ on p_f is also investigated. In all scenarios, p_f clearly increases with increasing correlation (from negative to positive) (Fig. 12). The maximum increase of p_f with increasing $\rho_{c',\varphi'}$ is of approximately 10% in scenarios 1, 2 and 3. The trend previously observed is not generally valid for scenario 4 (Fig. 12d): when p_f is

higher than 50% p_f decreases with increasing $\rho_{c',\varphi'}$. The reduction in p_f at time steps 2 and 3 days is equal to 6.33% and 3.69%, respectively.

Figure 13 shows this trend in terms of failure and safety regions, which are represented by 10^5 data points generated with a direct Monte Carlo simulation. The left side illustrates the results of scenario 2 at the time step 8 days, while scenario 4, at the time step 2 days (when p_f is higher than 50%), is shown on the right. A correlation between the shear parameters determines a deformation of the shape of the $c' - \varphi'$ region that includes all the realizations (green and red data points). Low and Tang [31] performed a probabilistic slope stability analysis with FORM and showed that two correlated normal variables in the original space form a tilted ellipsoid centered at the mean. A negative correlation implies more frequent combinations of low values of effective cohesion coupled with high values of effective friction angle and vice versa. When a positive correlation is considered, the region is tilted in the opposite direction, corresponding to more frequent combinations of high values of both effective cohesion and effective friction angle (or low values of both parameters). In scenarios 1, 2 and 3, when p_f is always lower than 50%, an increase in $\rho_{c',\varphi'}$ (from negative to positive) produces an increase in p_f . Conversely, in scenario 4 (time step 2 and 3 days), an increase in $\rho_{c',\varphi'}$ contributes to a reduction in the failure region. These results generally agree with those obtained by Griffiths et al. [13], where the authors state that when $p_f < 50\%$ a positive correlation is conservative and when $p_f > 50\%$ the opposite is true.

4.4 Effect of COV of the hydraulic parameters on p_f

The influence of the uncertainty of the hydraulic parameters k_s , a and n on p_f is examined. The hydraulic conductivity has a negligible influence on p_f , with a variation of p_f below 0.5% in all scenarios. Therefore, a positive or negative correlation between p_f and COV_{k_s} is not appreciable.

The parameter a influences p_f with an interesting tendency. The probability of failure increases with increasing COV_a (Fig. 14), but this variation becomes more significant when the minimum water level of 10 m is reached. The right side of Fig. 14 shows the variation of p_f with COV_a . The maximum variation of p_f observed in the 4 scenarios (ordered with increasing drawdown rate) is equal to 2.11%, 2.04%, 2.85% and 2.03%, respectively.

The influence of the curve fitting parameter n on p_f is very limited (Fig. 15). A slight increase of p_f with increasing COV_n is found before the minimum level of the reservoir is reached. This moment corresponds to the time steps 11, 5, 3 and 2 days in scenarios 1, 2, 3 and 4,

respectively. Afterwards, this tendency changes into a very small decrease of p_f with COV_n , below 0.6%.

4.5 Effect of the correlation of the curve fitting parameters a and n

The influence of the correlation between a and n is not significantly affecting p_f . For each scenario there is a slight tendency of p_f to decrease with increasing $\rho_{a,n}$ (from negative to positive), but the variation is lower than 1%. Evidence of this trend is provided by Wang et al. [51] and Jiang et al. [22].

5 Conclusions

This study proposes an intelligent surrogate model to estimate the probability of failure of reservoir slopes under rapid drawdown. This approach enables to perform reliability analysis of complex geotechnical problems with little computational effort. The LHS method is used to produce a sample of the soil parameters that constitutes the input features of the surrogate model. A deterministic approach that includes a transient unsaturated seepage and a slope stability analysis is applied for each sample combination to calculate FOS, which is the output predicted by the surrogate model. This dataset represents the training set of the surrogate model, which is then tested on a much larger sample, generated with a direct Monte Carlo simulation to accurately estimate the probability of failure. Two algorithms are tested and compared: SVR and XGBoost. XGBoost proves to be superior both in terms of computational efficiency and accuracy, with a testing performance in terms of R^2 above 98% against 94% of SVR. This model is employed to estimate the probability of failure of

a slope in four different scenarios with increasing draw-down rates.

A sensitivity analysis has provided insights into the influence of each soil parameter considered as stochastic (i.e., c' , φ' , k_s , a and n) on the probability of failure. The effective friction angle has, in terms of $COV_{\varphi'}$, the highest impact on p_f , followed by $COV_{c'}$, COV_a , COV_n and COV_{k_s} . The variation of p_f due to the COV of the random variables depends on the drawdown rate as well as the reservoir level. For a $COV_{\varphi'}$ growing from 0.05 to 0.20, p_f achieves an increase between 21% and 24% depending on the drawdown rate. It is also found that an increase in the COV generally causes a rise of p_f when this is lower than 50% (i.e., $FOS > 1$), but this effect is reversed when p_f is higher than 50%. This is caused by the position of the mean value of the shear strength parameters relative to the failure and safety regions. Furthermore, a positive correlation between the shear strength parameters c' and φ' is conservative when p_f is lower than 50%. A rise in p_f between 8 and 10% is observed in all scenarios when $\rho_{c',\varphi'}$ increases from -0.5 to 0.5 . A feature importance analysis has highlighted on which input parameters the surrogate model mostly depends, namely φ' and c' , with a maximum relative importance of 92% and 23%, respectively, depending on the drawdown rate and time step of the analysis. These results are consistent with those of the sensitivity analysis suggesting that reducing the variability of the effective friction angle with extensive site investigations would also aid in reducing p_f .

This work presents some limitations. The limit equilibrium method is applied to solve the slope stability analysis, and thus, the stress–strain behavior of the soil is not considered. Another issue is the spatial variability of soils, whose effect on p_f has not been investigated. However, this work represents a benchmark study of a time-dependent reliability analysis with intelligent surrogate models and a

basis of future research on the effect of the spatial variability of soils. Finally, this study has shown that intelligent surrogate models can be successfully used to perform probabilistic slope stability analysis and are capable of identifying the soil parameters with the major impact on the probability of failure.

Appendix

See Figs. 16, 17 and 18.

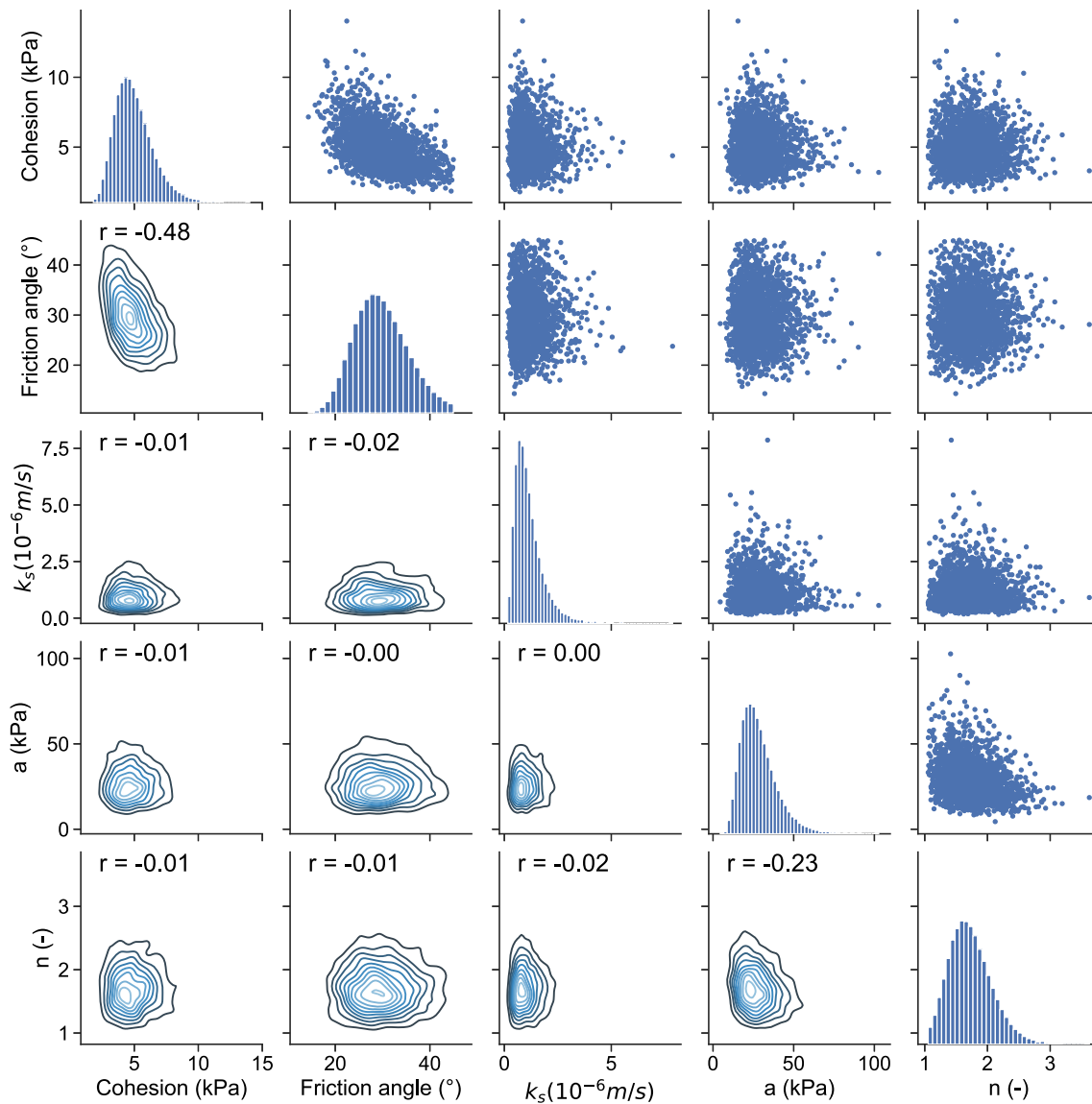


Fig. 16 Sample distribution and correlation between parameters

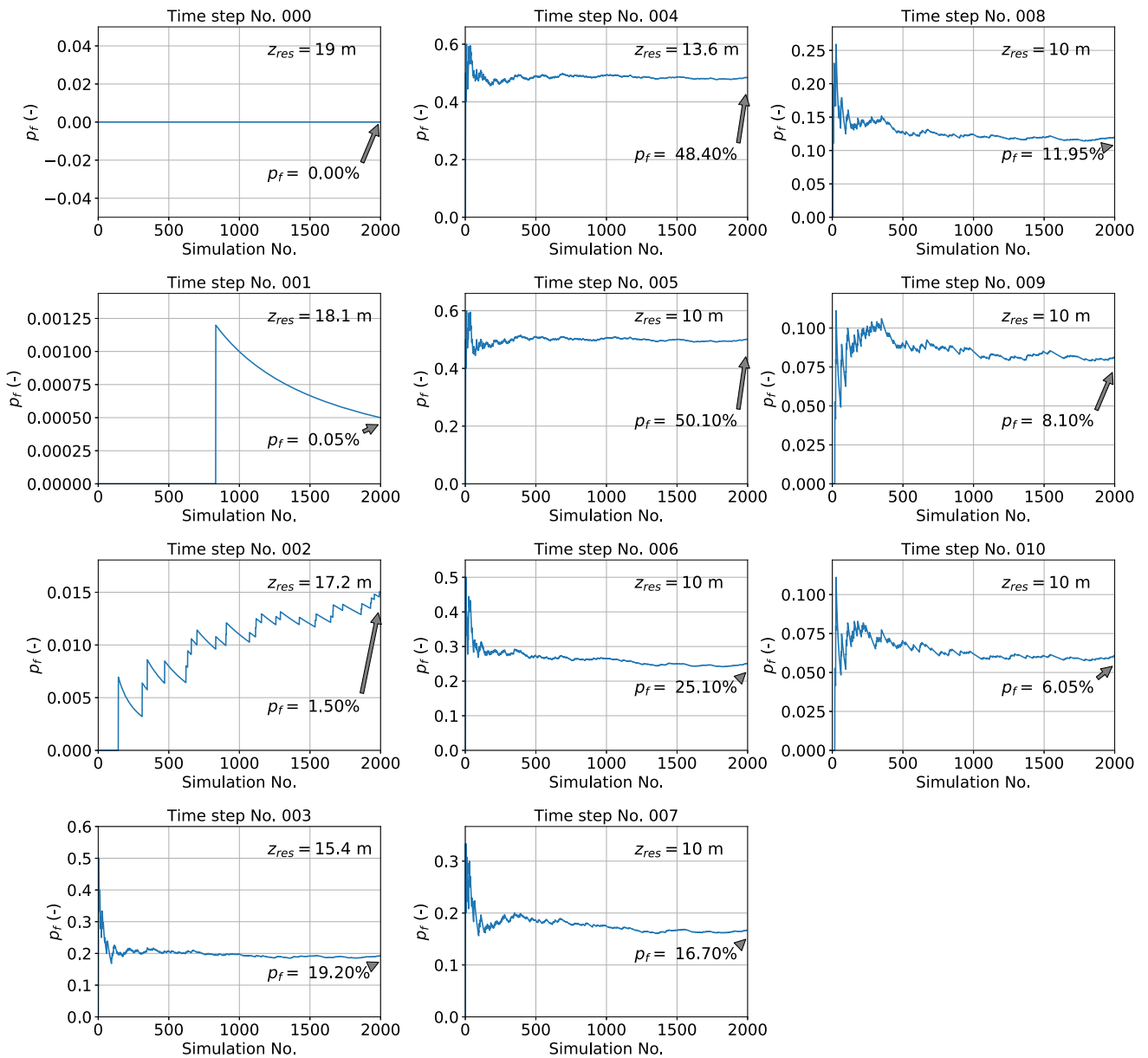


Fig. 17 Probability of failure estimated with GeoStudio (z_{res} is the reservoir level at the corresponding time step)

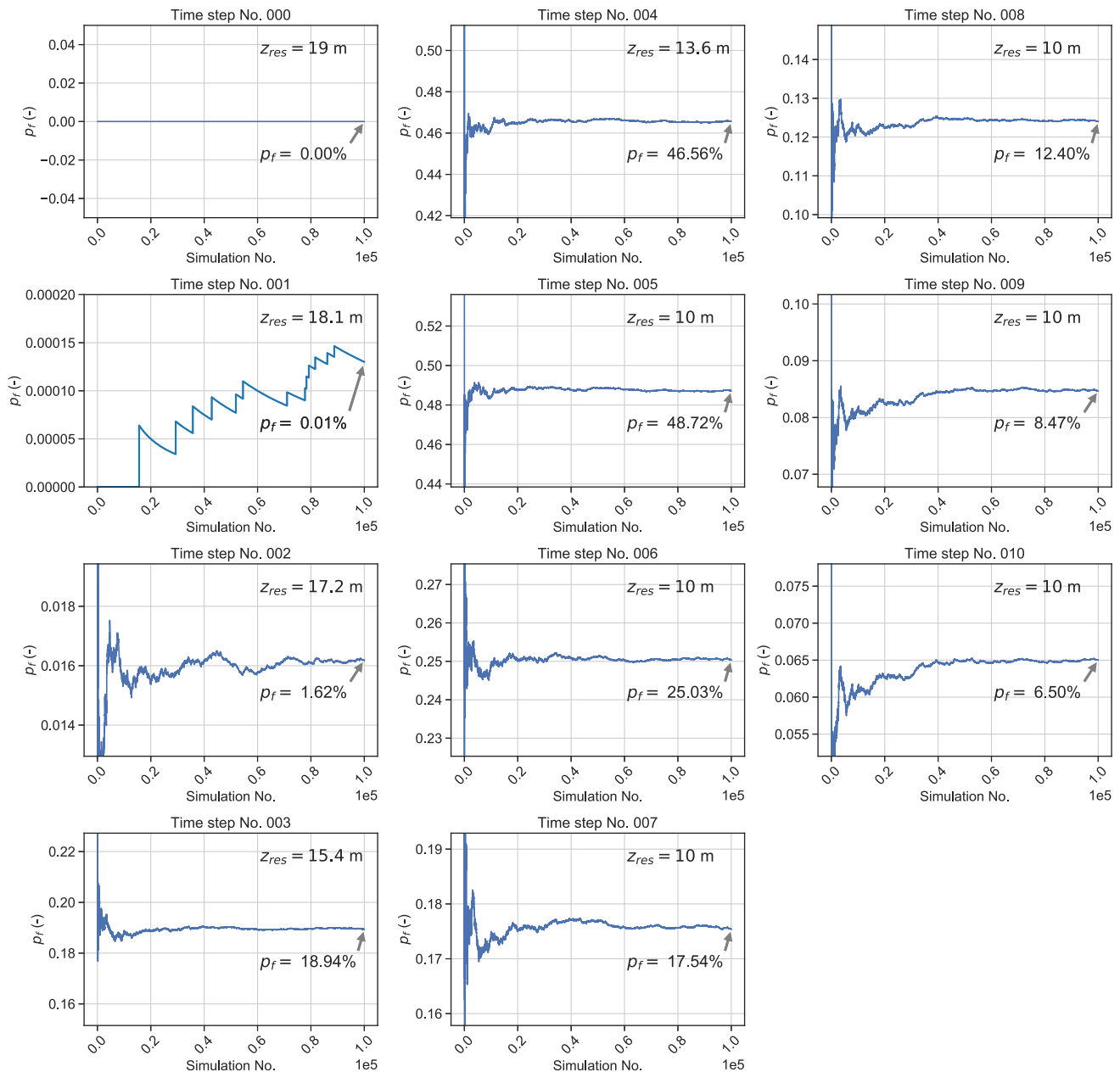


Fig. 18 Probability of failure estimated with XGBoost (z_{res} is the reservoir level at the corresponding time step)

Acknowledgements Financial support for this research is provided by the Otto Pregl Foundation for Geotechnical Fundamental Research and the H2020 Marie Skłodowska-Curie Actions RISE 2017 “HERCULES” with Grant No. 778360.

Funding Open access funding provided by University of Natural Resources and Life Sciences Vienna (BOKU).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonso EE (1976) Risk analysis of slopes and its application to slopes in Canadian sensitive clays. *Géotechnique* 26(3):453–472. <https://doi.org/10.1680/geot.1976.26.3.453>

2. Ang AHS, Tang WH (2006) Probability concepts in engineering: emphasis on applications to civil and environmental engineering, 2nd edn. Wiley, New York
3. Awad M, Khanna R (2015) Support vector regression. Apress, Berkeley. https://doi.org/10.1007/978-1-4302-5990-9_4
4. Bojer CS, Meldgaard JP (2021) Kaggle forecasting competitions: an overlooked learning opportunity. *Int J Forecast* 37(2):587–603. <https://doi.org/10.1016/j.ijforecast.2020.07.007>
5. CEN (European Committee for Standardization) (2002) Eurocode—basis of structural design. Standard, Brussels
6. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, association for computing machinery. New York, NY, USA, KDD'16, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
7. Christian JT, Ladd CC, Baecher GB (1994) Reliability applied to slope stability analysis. *J Geotechn Eng* 120(12):2180–2207. [https://doi.org/10.1061/\(asce\)0733-9410\(1994\)120:12\(2180\)](https://doi.org/10.1061/(asce)0733-9410(1994)120:12(2180))
8. Deutsch JL, Deutsch CV (2012) Latin hypercube sampling with multidimensional uniformity. *J Stat Plann Inference* 142(3):763–772. <https://doi.org/10.1016/j.jspi.2011.09.016>
9. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines. *Adv Neural Inform Process Syst* 9:155–161
10. Gao X, Liu H, Zhang W, Wang W, Wang Z (2018) Influences of reservoir water level drawdown on slope stability and reliability analysis. *Georisk: Assess Manag Risk Eng Syst Geohazards* 13(2):145–153. <https://doi.org/10.1080/17499518.2018.1516293>
11. Genuchten M (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44(5):892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
12. Griffiths DV, Fenton GA (2004) Probabilistic slope stability analysis by finite elements. *J Geotechn Geoenviron Eng* 130(5):507–518. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2004\)130:5\(507\)](https://doi.org/10.1061/(ASCE)1090-0241(2004)130:5(507))
13. Griffiths DV, Huang J, Fenton GA (2009) Influence of spatial variability on slope reliability using 2-D random fields. *J Geotechn Geoenviron Eng* 135(10):1367–1378. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000099](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000099)
14. Head T, Kumar M, Nahrstaedt H, Louppe G, Shcherbatyi I (2020) Scikit-optimize/scikit-optimize. <https://doi.org/10.5281/zenodo.4014775>
15. Huang J, Fenton G, Griffiths DV, Li D, Zhou C (2017) On the efficient estimation of small failure probability in slopes. *Landslides* 14(2):491–498. <https://doi.org/10.1007/s10346-016-0726-2>
16. ICOLD (2016) Small dams: design, Surveillance and Rehabilitation. Technical Report Bulletin 157, International Commission on Large Dams
17. Iman RL, Conover WJ (1982) A distribution-free approach to inducing rank correlation among input variables. *Commun Stat Simul Comput* 11(3):311–334. <https://doi.org/10.1080/03610918208812265>
18. Javankhoshdel S, Bathurst RJ (2014) Simplified probabilistic slope stability design charts for cohesive and cohesive-frictional (c- ϕ) soils. *Can Geotech J* 51:1033–1045. <https://doi.org/10.1139/cgj-2013-0385>
19. Jiang SH, Huang JS (2016) Efficient slope reliability analysis at low-probability levels in spatially variable soils. *Comput Geotech* 75:18–27. <https://doi.org/10.1016/j.compgeo.2016.01.016>
20. Jiang SH, Li DQ, Zhang LM, Zhou CB (2014) Slope reliability analysis considering spatially variable shear strength parameters using a non-intrusive stochastic finite element method. *Eng Geol* 168:120–128. <https://doi.org/10.1016/j.enggeo.2013.11.006>
21. Jiang SH, Li DQ, Cao ZJ, Zhou CB, Phoon KK (2015) Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. *J Geotechn Geoenviron Eng* 141(2):04014096. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001227](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001227)
22. Jiang SH, Liu X, Huang J (2020) Non-intrusive reliability analysis of unsaturated embankment slopes accounting for spatial variabilities of soil hydraulic and shear strength parameters. *Eng Comput*. <https://doi.org/10.1007/s00366-020-01108-6>
23. Kang F, Li J (2016) Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes. *J Comput Civ Eng* 30(3):04015040. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000514](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000514)
24. Kang F, Han S, Salgado R, Li J (2015) System probabilistic stability analysis of soil slopes using Gaussian process regression with Latin hypercube sampling. *Comput Geotech* 63:13–25. <https://doi.org/10.1016/j.compgeo.2014.08.010>
25. Kang F, Li J, Li J (2016a) System reliability analysis of slopes using least squares support vector machines with particle swarm optimization. *Neurocomputing* 209:46–56. <https://doi.org/10.1016/j.neucom.2015.11.122>
26. Kang F, Xu Q, Li J (2016b) Slope reliability analysis using surrogate models via new support vector machines with swarm intelligence. *Appl Math Model* 40(11):6105–6120. <https://doi.org/10.1016/j.apm.2016.01.050>
27. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International joint conference on artificial intelligence (IJCAI), Montreal, Canada, vol 14, pp 1137–1145
28. Lane PA, Griffiths DV (2000) Assessment of stability of slopes under drawdown conditions. *J Geotechn Geoenviron Eng* 126(5):443–450. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2000\)126:5\(443\)](https://doi.org/10.1061/(ASCE)1090-0241(2000)126:5(443))
29. Li S, Zhao HB, Ru Z (2013) Slope reliability analysis by updated support vector machine and Monte Carlo simulation. *Nat Hazards* 65(1):707–722. <https://doi.org/10.1007/s11069-012-0396-x>
30. Liao K, Wu Y, Miao F, Li L, Xue Y (2021) Time-varying reliability analysis of Majiagou landslide based on weakening of hydro-fluctuation belt under wetting-drying cycles. *Landslides* 18(1):267–280. <https://doi.org/10.1007/s10346-020-01496-2>
31. Low BK, Tang WH (1997) Probabilistic slope analysis using Janbu's generalized procedure of slices. *Comput Geotech* 21(2):121–142. [https://doi.org/10.1016/S0266-352X\(97\)00019-0](https://doi.org/10.1016/S0266-352X(97)00019-0)
32. Ltd GSI (2018) Geostudio. <https://www.geoslope.com/>
33. Millard SP (2013) EnvStats: an R Package for environmental statistics, 2nd edn. Springer, New York. <https://doi.org/10.1007/978-1-4614-8456-1>
34. Morgenstern N (1963) Stability charts for earth slopes during rapid drawdown. *Géotechnique*. <https://doi.org/10.1680/geot.1963.13.2.121>
35. Moza S (2020) sahil89/lhsmdu: Latin Hypercube Sampling with Multi-Dimensional Uniformity (LHSMU): Speed Boost minor compatibility fixes. <https://doi.org/10.5281/zenodo.3929531>
36. Mualem Y (1976) A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour Res* 12(3):513–522. <https://doi.org/10.1029/WR012i003p00513>
37. Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Le HV, Tran VQ, Prakash I, Pham BT (2021) Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Math Probl Eng*. <https://doi.org/10.1155/2021/4832864>
38. Papaioannou I, Betz W, Zwirgmaier K, Straub D (2015) MCMC algorithms for Subset Simulation. *Probab Eng Mech* 41:89–103. <https://doi.org/10.1016/j.probgemch.2015.06.006>
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,

- Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12(85):2825–2830
40. Phoon KK, Ching J (2015) Risk and reliability in geotechnical engineering. CRC Press, Boca Raton. <https://doi.org/10.1201/b17970>
41. Phoon KK, Santoso A, Quek ST (2010) Probabilistic analysis of soil-water characteristic curves. *J Geotechn Geoenviron Eng* 136(3):445–455. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000222](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000222)
42. Pinyol NM, Alonso EE, Olivella S (2008) Rapid drawdown in slopes and embankments. *Water Resour Res*. <https://doi.org/10.1029/2007WR006525>
43. Pinyol NM, Alonso EE, Corominas J, Moya J (2012) Canelles landslide: modelling rapid drawdown and fast potential sliding. *Landslides* 9(1):33–51. <https://doi.org/10.1007/s10346-011-0264-x>
44. Richards LA (1931) Capillary conduction of liquids through porous mediums. *Physics* 1(5):318–333. <https://doi.org/10.1063/1.1745010>
45. Siacara AT, Beck AT, Futai MM (2020) Reliability analysis of rapid drawdown of an earth dam using direct coupling. *Comput Geotech* 118:103336. <https://doi.org/10.1016/j.compgeo.2019.103336>
46. Suthaharan S (2016) Machine learning models and algorithms for big data classification: thinking with examples for effective learning. Integrated series in information systems. Springer, USA. <https://doi.org/10.1007/978-1-4899-7641-3>
47. Vanapalli SK, Fredlund DG, Pufahl DE, Clifton AW (1996) Model for the prediction of shear strength with respect to soil suction. *Can Geotech J* 33(3):379–392. <https://doi.org/10.1139/t96-060>
48. Vorechovský M, Novák D (2009) Correlation control in small-sample Monte Carlo type simulations I: a simulated annealing approach. *Probab Eng Mech* 24(3):452–462. <https://doi.org/10.1016/j.probingmech.2009.01.004>
49. Wang ZZ, Goh SH (2021) Novel approach to efficient slope reliability analysis in spatially variable soils. *Eng Geol* 281:105989. <https://doi.org/10.1016/j.enggeo.2020.105989>
50. Wang Y, Huang J, Tang H (2019) Global sensitivity analysis of the hydraulic parameters of the reservoir colluvial landslides in the Three Gorges Reservoir area, China. *Landslides* 17(2):483–494. <https://doi.org/10.1007/s10346-019-01290-9>
51. Wang L, Wu C, Gu X, Liu H, Mei G, Zhang W (2020a) Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. *Bull Eng Geol Environ* 79(6):2763–2775. <https://doi.org/10.1007/s10064-020-01730-0>
52. Wang L, Wu C, Tang L, Zhang W, Lacasse S, Liu H, Gao L (2020b) Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. *Acta Geotech* 15(11):3135–3150. <https://doi.org/10.1007/s11440-020-00962-4>
53. Wu Y, Miao F, Li L, Xie Y, Chang B (2017) Time-varying reliability analysis of Huangtupo Riverside No.2 Landslide in the Three Gorges Reservoir based on water-soil coupling. *Eng Geol* 226:267–276. <https://doi.org/10.1016/j.enggeo.2017.06.016>
54. Zhang W (2020) MARS applications in geotechnical engineering systems: multi-dimension with big data. Springer, Singapore. <https://doi.org/10.1007/978-981-13-7422-7>
55. Zhang J, Huang HW, Phoon KK (2013) Application of the kriging-based response surface method to the system reliability of soil slopes. *J Geotechn Geoenviron Eng* 139(4):651–655. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000801](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000801)
56. Zhang W, Tang L, Li H, Wang L, Cheng L, Zhou T, Chen X (2020) Probabilistic stability analysis of Bazimen landslide with monitored rainfall data and water level fluctuations in Three Gorges Reservoir, China. *Front Struct Civ Eng* 14(5):1247–1261. <https://doi.org/10.1007/s11709-020-0655-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.