Risk analysis of infinite slope failure using advanced Bayesian networks

Longxue He[1], António Topa Gomes[2], Matteo Broggi[1], and Michael Beer[1,3,4*]

[1] *Institute for Risk and Reliability, Leibniz Universität Hannover, Germany*
[2] *CONSTRUCT, Faculty of Engineering, University of Porto, Portugal*
[3] *Institute for Risk and Uncertainty, University of Liverpool, UK*
[4] *CERSM, Tongji University, China*
[*] *Corresponding author: beer@irz.uni-hannover.de*

Abstract

Slope stability is affected by multiple factors. To assess the risk of slope failure, it is critical to identify the key influence factors and accurately predict the probability of slope failure. For that purpose, Bayesian networks (BNs) are an effective tool. This paper presents the application of advanced BNs to evaluate the geotechnical risk of slopes. The first part of the paper presents the theoretical background regarding BNs. It includes the simple BNs, and explains further developments. Particularly, enhanced Bayesian networks, able to cope with continuous input parameters, and Credal networks, specially used for incomplete input information, are presented and discussed. The second part of the paper deals with two geotechnical examples implemented to demonstrate the feasibility and predictive effectiveness of BNs. The developments regarding the use of BNs in geotechnical engineering are presented. The ability of BNs to deal with the slope stability risk is discussed as well. The paper also evaluates the influence of several geotechnical parameters. Besides, it discusses how the different types of BNs contribute to assessing the reliability and risk of real slopes, and how new information could be introduced in the analysis.

Keywords: Risk analysis, Infinite slope, enhanced Bayesian networks, Credal networks

1 Introduction

The occurrence of slope failure has the catastrophic potential for causing landslides such as debris flows and rock falls, which threatens the human life and property, attracting public attention and requiring research in geotechnical engineering [1, 2]. Various attempts have been carried out in literature to analyze the stability of slopes and estimate the probability of slope failure. In these approaches, infinite slope model is widely utilized to study shallow landslides under the varying position of the groundwater table, random variations of soil parameters and so on [3–5].

The probabilistic method has played an important role in the estimation of the probability of failure of slopes since the very early studies [6–10] due to the unavoidable uncertainties existing in vague environmental condition, varying soil properties as well as insufficient information affecting the slope failure.

Numerous numerical methods have been applied in slope stability analysis. For instance, slope stability problems associated with structural reliability methods (SRMs) have been conducted by means of first-order reliability methods (FORMs) [8] and simulation approaches, such as Monte Carlo Simulation [11], Importance Sampling [12] and Subset sampling [13], etc. These studies demonstrated the feasibility of structural reliability analysis for calculating the probability of slope failure in geotechnical engineering. However, the interactive influences of the multiple parameters in the analysis of a slope failure cannot be identified and further singled out the critical factors with these approaches.

Artificial neural networks also have been adopted to predict slope stability of slopes with the geometric or geological data of the influential factors [14, 15]. This approach is not good at quantifying the uncertainty and characterizing the impact of individual risk factors on the slope stability using information updating.

As is known, the identification of the most important factors leading to failure and the estimation of the probability of slope failure can prevent potential geological disasters. A robust model presented in recent researches, Bayesian networks (BNs) have been increasingly employed in slope stability analysis [16–20]. BNs, as the causal probabilistic models, have been developed and successfully applied to natural hazards, safety, and reliability engineering for over two decades since their first introduction by Pearl [21]. Compared to the aforementioned numerical tools, BNs carries huge advantages over other available methods to calculate the probability of slope failure. In particular, they show the following advantages.

- Simple graphical visualization. The failure of a slope can be affected by geo-environmental parameters, weather condition, natural hazards (e.g. earthquakes and storm) as well as human activities. The combination of these factors can negatively impact on slope stability. BNs can not only integrate these elements into a rigorous framework but provide a visual cause-effect relation among events in a graphical model. In particular, BNs help decision makers and even non-expert without a strong background in geotechnical engineering to gain a good understanding of the failure mechanisms. For a detailed overview on how to construct a graphical framework for risk assessment of rock-fall hazard with a BN model is given by [22].

- Uncertainty quantification. BNs are developed successfully to capture the uncertainties affecting the problem and benefit from the capability of the forward and backward propagation of probabilities according to the axioms of Bayesian probability theory [23].

- Information update from new observation. Updating of the event probabilities in BNs can be efficiently performed in near-real-time by mean of Bayesian updating to respect the information carried by the new observation. Thanks to this, the BNs model can provide the decision makers with up-to-date information on the slope failure mechanisms as soon as new evidence is presented.

In previous studies, traditional BNs (i.e., only discrete probability values and binary event are considered) have been applied to analyse slope stability. However, the slope stability problem is clearly influenced by both discrete events and continuous variables, thus it is impractical to obtain discrete probabilities of all the factors affecting a slope. Thanks to the development of enhanced Bayesian networks (eBNs), proposed by Straub & Der Kiureghian [24], it is possible to deal also with a causal relationship to continuous quantities in a BN framework.

An additional limitation of BNs is that the employed probabilistic models are precise, hindering the application in geotechnical engineering where the available information is often scarce. Thus, Credal networks (CNs), an extension of BNs to take imprecise probabilities into account , have been introduced in the geotechnical problem to assess debris flow hazard [25]. Nonetheless, no detailed investigations are attempted to apply the advanced model into the analysis of slope stability.

In this paper, the study aims at presenting the graphical models of the slope to estimate the failure probability, in which both of eBNs and CNs implemented are based on the infinite slope, and then new observations are inserted to update the model in order to identify the effect of factors on slope stability.

2 Methodology

2.1 Bayesian Networks

Bayesian networks, also known as Bayesian belief networks or causal networks, originate from artificial intelligence and statistics [21, 26]. They were developed as a powerful modeling tool for decision support and quantification of uncertainties, especially for low probability events. They have been applied to risk analysis in many studies since 2001 [27].

In a nutshell, a Bayesian network (see Figure 1) is a directed acyclic graph, in which a set of variables are represented by nodes. The relation between each node is represented in terms of parent-child and linked by an arrow, denoting the conditional dependencies between these variables. Conditional Probability Tables (CPTs) are attached to each node and consider all the possible states of a variable. Then, the probabilities of the nodes are determined by factorization of the joint probabilities by means of Bayes' theorem. The joint probability is the function of all the random variables in BNs. For any BN, it can be given mathematically by a product of the CPTs entries,

$$P(X_i) = \prod_{i}^{n} P(X_i|pa(X_i)) \tag{1}$$

where $X_i = X_1, \ldots, X_n$ denote the nodes of the BN, $pa(X_i)$ are the set of parents of X_i, and $P(X_i|pa(X_i))$ represent the entries of the CPTs. The effective methods for general inference in BNs can be achieved in [28], and it is also applicable for probability updating. For instance, in the case where evidence is assigned to the observed nodes $X_j = e$, this information will propagate through the prior probabilities to the posterior probabilities as follows,

$$P(X_i|e) = \frac{P(X_i, e)}{P(e)} = \frac{\prod_{i}^{n} P(X_i|pa(X_i), e)}{\sum_{X_i \setminus X_j} P(X_i, e)}. \tag{2}$$

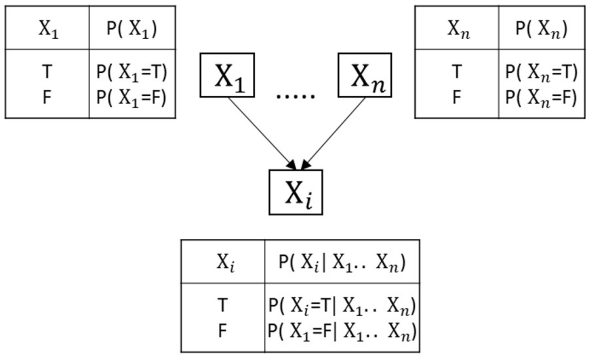

Figure 1: A simple graph of a general BN (T=True; F=False).

note that the joint distribution $P(X_i, e)$, obtained by using equation (1), associates with the evidence value e, and compute $P(e)$ from $P(X_i, e)$ by marginalizing out all the variables except the node X_j. If a node with no children has no associated evidence, it is called "barren node" meaning that the conditional probability is useless for the calculation of the marginal probabilities of non-barren nodes [29].

In general, as for the ability of belief propagation in the network, marginal posterior probabilities of the query nodes can be achieved through both top to bottom and inverse reasoning by means of the inference algorithms [21, 30], including exact algorithms and approximate algorithms. In comparison to approximate algorithms, exact algorithms, which are suitable for computing discrete BNs, are guaranteed to gain correct answers and hence it is a more robust computational method. In case of continuous variables in a BN, however, given the difficulty of defining the prior probability distributions as the discrete form, unavoidably impeding the application of BNs to the practical field.

BNs consisting of discrete and continuous variables are referred to as hybrid BNs. With consideration of exact algorithms, there are three special approaches to extending discrete BNs to continuous BNs or hybrid BNs. The first is to restrict continuous nodes to Gaussian random variables while allowing them to link only towards their non-discrete children. The second method is to define the continuous nodes as a mixture of truncated exponential distributions (MTEs), which is a generalization allowing to approximate any distribution function, but still requires further scrutiny [31]. The final methodology is enhanced Bayesian networks, implemented by joining BNs with SRMs, and was successfully applied in risk and reliability analysis [32, 33]. An introduction to this method is given in details in the following section.

2.2 Enhanced Bayesian Networks

In a system reliability problem, the outcome domain of an event, determined by a set of continuous random variables with known distributions, can be divided into failure and safe region by the relevant limit state functions. The failure probability of an event is the integral of the probability density function in the failure domain. This is a very complex integral and in general intractable analytically. Its value can be approximated by means of first or second order reliability methods or simulation approaches. Enhanced BNs approach is to combine SRMs with BNs. Specifically, in a BN, the continuous nodes must have at least an offspring, which is a discrete node defined as a domain in the outcome space of these continuous nodes. That is, the continuous nodes should meet the requirement of well-established SRMs, and it is the key condition for using eBNs approach.

Then these continuous nodes lose the causal dependencies with their deterministic nodes by means of FORM and SORM or simulation approaches. Finally, hence, all the continuous nodes can be removed from eBNs according to node elimination algorithm [24], and thus hybrid BNs are reduced to discrete BNs.

An example of computation of the total probability of an eBN and the process of node elimination is described by equation (3) to equation (5) for the simple case represented by Figure 2. From of equation (1), the joint probability of all the nodes for the eBN can be written as

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_4|X_2, X_3)f(X_3)f(X_2|X_1) \tag{3}$$

in which $P(X_1)$ and $P(X_4|X_2, X_3)$ represent conditional probabilities of discrete nodes X_1 and X_4 while $f(X_3)$ and $f(X_2|X_1)$ are the probability density functions of continuous nodes X_2 and X_3, respectively.

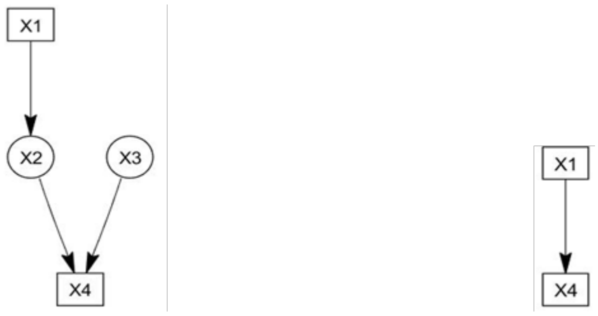

Figure 2: An example of reduction of an eBN into BN.

The joint probability of the discrete nodes can be obtained by marginalization calculation of equation (3) as:

$$P(X_1, X_4) = \int\int_{X_2,X_3} P(X_1)P(X_4|X_2, X_3)f(X_3)f(X_2|X_1)\, dX_2\, dX_3 \tag{4}$$

in the case the domain of node X_4 can be determined by the outcome space of its parent nodes, then equation (4) can be rearranged as:

$$P(X_1, X_4) = P(X_1) \int\int_{\Omega_{X_4}(X_2,X_3)} f(X_3)f(X_2|X_1)\, dX_2\, dX_3 \tag{5}$$

where $\Omega_{X_4}(X_2, X_3)$ represents variable X_4 as a domain in the outcome space of variables X_2 and X_3. The form of equation (5) are in line with the definition of structural reliability problems, and hence can be estimated by means of SRMs.

2.3 Evidence on Continuous nodes

As already stated, BNs show a powerful capability in updating probabilistic propagation through given observations. As previously discussed, evidence is inserted to replace certain priori probability on observed nodes, and the probabilities of the other nodes are updated using exact algorithms in discrete BNs. In a similar way, in eBNs, it is necessary to discretize continuous nodes with evidence at first, and then the corresponding discrete nodes are kept in place of the continuous nodes in the reduced BNs.

A plethora of discretization methods for continuous nodes in the BNs has been being investigated in [31, 34–37]. Currently, there are no formalized approaches for discretization of continuous random variables. Thus, in terms of the study problem here, a credible discretization approach for eBNs [24] is used in this study.

In terms of the study problem in this paper, only discretization of continuous nodes without any parent is introduced here. The previously introduced example is here reintroduced to explain how to discretize continuous nodes in eBNs. As shown in Figure 3, node X_3 is substituted with two nodes, a discrete variable $X3_{discrete}$ and a continuous variable $X3_{continuous}$.

$X3_{discrete}$ has i states that are defined by the outcome space of X_3 with conditional cumulative distribution function $F_{X_3}[x_3]$, and the number of its states is identical to corresponding intervals of the divided domain of X_3. Each sub-domain of X_3 can be represent by $[\underline{x}_{3i}, \overline{x}_{3i}]$, where $\underline{x}_{3i}$ and $\overline{x}_{3i}$ denote the lower and upper bounds of the interval, respectively. Then the probability mass function of $X3_{discrete}$ given the state i can be achieved as,

$$P(X^i_{3\,discrete}) = F_{X_3}\left[\overline{x_{3i}}\right] - F_{X_3}\left[\underline{x_{3i}}\right] \tag{6}$$

On the other hand, $X3_{continuous}$, as the child of $X3_{discrete}$, inherits all the descendants and outcome space of X_3. The continuous variable $X3_{continuous}$ is eliminated from the model after it becomes a barren node by used of SRMs, and the discretized node $X3_{discrete}$ is retained to facilitate new observations updating the model.

In the same way, for inserting the evidence on X_3, the process of discretization is to split the domain of X_3 given the evidence into the sub-domains, each of which is obtained with a discrete probability value. In this study, the number of the sub-domains on the observed continuous node is defined by five with the same interval.

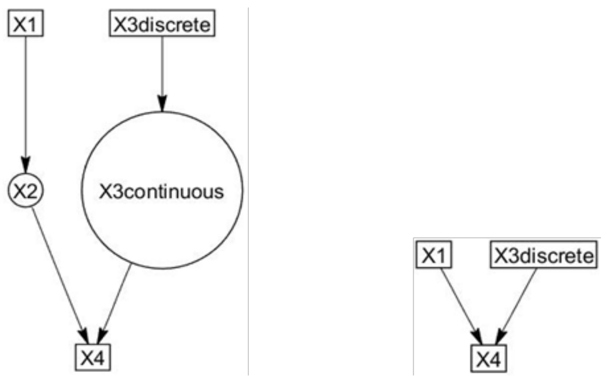

Figure 3: An example of discretization procedure.

2.4 Credal Networks

In the case imprecise probabilities are introduced to Bayesian networks, they are referred as Credal networks (CNs) since the node corresponding to an imprecise event is associated with a credal set instead of a CPT or PDF. Credal sets are defined as closed convex sets associated with a set of probability distribution functions, which are used to represent imprecise probabilities in the graphical models. Fagiuoli and Zaffalon [38] used convex sets to compute posterior probabilities in a discrete BN with exact algorithms, and first referred to this kind of model as Credal networks. A detailed introduction of credal networks can be found in the literature [39, 40].

The inference in CNs is more complex than BNs, which is studied by some researchers [41–45]. Thanks to the development of inference algorithms in CNs, some exact and approximate inference algorithms can be used for the reasoning of CNs although imprecise probabilities propagation in CNs is still under study. In this paper, an integration of CNs and SRMs [46] is adopted to analyse the slope problems.

2.4.1 Inference computation in CNs

Continuous variables and interval variables in CNs are computed in terms of eBNs approach. The computation condition and elimination procedure (Figure 4) is the same with eBNs. After removing the continuous and imprecise nodes, CNs only contain two types of conditional probability in discrete nodes: point probabilities and bounded probabilities. Afterwards, variable elimination algorithm, as a classical exact inference algorithm, is applied here to estimate probability propagation in CNs.

As an example, a simple Credal network(CN) is shown in Figure 4. It consists of three types of nodes: discrete node X_1, continuous node with a known distribution X_2 , and an imprecise node X_3. The deterministic node X_4 is dependent of all the other three nodes. Besides, both of discrete nodes X_1 and X_4 belong to binary variables, and then the joint probability for identifying upper and lower bounds of nodes in the CN can be expressed as,

$$P(X_1, \overline{\underline{X}}_4) = P(X_1)P(\overline{\underline{X}}_4|X_1) \tag{7}$$

in which $\overline{\underline{X}}_4$ denotes the upper and lower bounds in node X_4 with two states x_{41} and x_{42} Then, according to variable elimination, exact bounds of marginal probability with upper bound in the state x_{11} of node X_1 can be obtained as,

$$P(\overline{x}_{11})_{exact} = \max\left(\sum_{\overline{\underline{X}}_4} P(X_1)P\left(\overline{\underline{X}}_4|X_1\right)\right) = \max\begin{bmatrix} P(x_{11})P(\underline{x}_{41}|x_{11}) + P(x_{11})P(\overline{x}_{42}|x_{11}) \\ P(x_{11})P(\overline{x}_{41}|x_{11}) + P(x_{11})P(\underline{x}_{42}|x_{11}) \end{bmatrix} \tag{8}$$

The lower bound of the marginal probability can be obtained in a similar way with the minimum operator. Although traditional exact inference algorithms are efficient to compute the exact bounds, the exact inference is highly inefficient and leads to a combinatorial explosion in the case of complex networks, since it requires the evaluation of every possible bound combination for every node.

A novel algorithm has been introduced to avoid this combinatorial explosion encountered by exact inference [47]. The outcome from this approach can get the inner bounds, which can be equal to the exact bounds if no nodes with probability interval are observed. For a query node, briefly, instead of computing the true bound with identifying all of the combinations of the bounds in input, the key step is to compare the conditional probabilities of the query variable given the related nodes in CNs. Therefore,

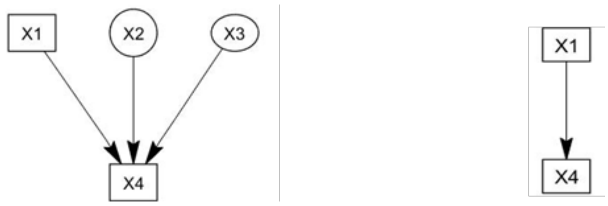

Figure 4: An example of elimination procedure in a simple CN.

it is obvious that the result by use of this kind of inner approximation is exact if no evidence involved in the bounded nodes.

In a word, this approach makes the computation low-cost, and it is effective to obtain real-time results with respect to the imprecise nodes in the model.

2.4.2 Simulation methods for the model with continuous or imprecise nodes

Slope stability analysis is commonly based on the evaluation of the factor of safety and involves a large number of random variables [48], and the probability of the slope failure can be obtained by means of SRMs with various simulation methods. Direct Monte Carlo approach is a robust and feasible method to compute the probability of failure. It has been previously applied to slope stability [8] and it is a classical simulation tool suited for the reduction of eBNs. Nevertheless, it requires a very high number of samples in the case of small failure and high dimensions. This is especially the case in the analysis of slope failures, where failure probabilities are typically in the order of 10^{-4} or smaller.

Advanced line sampling [49] is a recently developed advanced Monte Carlo methods, based on line sampling [50], and employs an adaptive algorithm to adapt the important direction to the shape of limitation state surface. Most importantly, it allows for sets of probability distributions to be included in the estimation of imprecise failure probabilities, which are bounded with upper and lower probabilities. Because of these advantages, advanced line sampling is adopted for the evaluation of CNs in this paper.

3 Risk assessment of slope stability with advanced BNs

The slope stability analysis considers the driving forces and resisting forces in the slope, comparing them and calculating the Factor of Safety. When the driving forces are larger than the resistant forces, the slope fails. The model introduced in the following section is to construct a cause-effect graph to estimate the probability of slope failure and identify the key landslide-induced factors.

3.1 Factors of safety for an infinite slope

Factors of safety are frequently computed to identify whether a slope is safe, which can be obtained by the ratio of resisting and driving stresses along a potential slip surface. This calculation, however, is not based on a unique equation, since there are a variety of methods [51–54] that can be selected to obtain the factor of safety according to the different conditions. These conditions also depend on the type of failure surface and its extension.

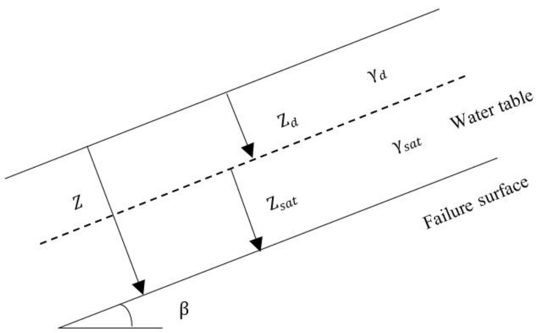

Figure 5: An infinite slope.

For an infinite slope, as the Figure 5 shown, the equation for the factor of safety in terms of effective stress analysis is given by

$$FOS = \frac{c + (\gamma_d Z_d + \gamma_{sat} Z_{sat} - \gamma_w Z_{sat}) \cdot \cos\beta \cdot \tan\phi}{(\gamma_d Z_d + \gamma_{sat} Z_{sat}) \cdot \sin\beta} \tag{9}$$

here, the drained parameters of cohesion (c) and friction angle (ϕ) are the main properties of soil. Z_d and Z_{sat} are the thickness of unsaturated and saturated soil layer, respectively, and the sum of them is the total thickness of soil (Z). β is the slope inclination and γ_w is the unit weight of water, 9.81 kN/m^3. For the layer above and below water table, soil unit weight should be split into two parts: dry unit soil weight (γ_d) and saturated unit soil weight (γ_{sat}). This analysis has been completed using in infinite slope stability model [55]. Moreover, $FOS \leq 1$ means the slope failure, whilst the FOS larger than 1 indicates the slope is safe. All the calculations are performed in effective stresses but, for the sake of simplicity, the effective parameters, cohesion and friction angle, are simply denominated as c and ϕ, as there is no risk to misunderstand effective and total strength resistances.

3.2 The stability analysis of an infinite slope with BNs

The slope BN (see Figure 6) considers the stability of an infinite slope with its relevant variables and factors of safety. Therefore, it is reasonable to take factors of safety as a target event to infer the crucial events affecting the consequence, and further ascertain the causal relation amongst the events.

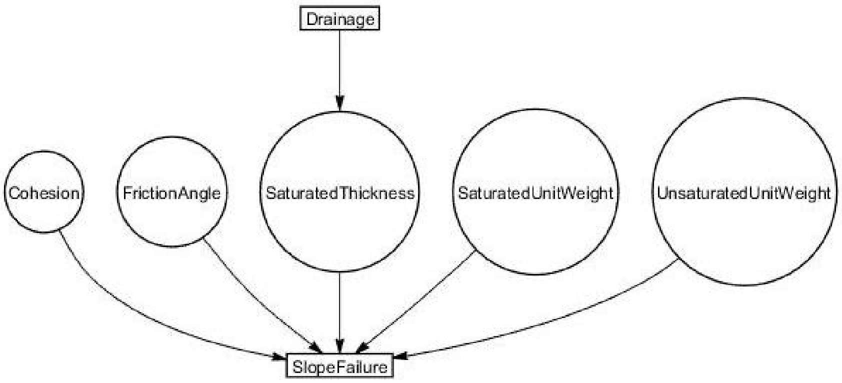

Figure 6: The hybrid BN model of an infinite slope.

The model includes six factors and one failure event. The main strength parameters of the soil, Cohesion and Friction Angle are the resisting forces preventing the occurrence of a failure. Meanwhile, the geometrical parameters of the slope are the slope inclination and slope's height, being also two important factors for slope stability. The angle of a slope defines how much driving force is distributed in the parallel direction along the slope surface. Small angles mean small pulling force on the downslope movement while large angle provides the large pulling force. In this model, Z and β are constant, and the height of slope can be obtained by $Z/\cos\beta$, so they are not considered into this BN.

According to effective stress principle [56], pore water pressure is defined by the unit weight of soil and the corresponding soil thickness. In such conditions, it was also considered the influence of the water table in the slope stability. In light of this, the nodes Unsaturated Unit Weight, Saturated Unit Weight and Saturated Thickness are selected in the slope model.

The position of groundwater table is an unstable variable defining the slope safety. The node Saturated Thickness represents the depth of saturated soil, which is the level of the water table. This random variable is governed by the drainage condition. To be specific, the water table is away when drainage takes place. If not, the depth of saturated soil will assume random values ranging between 0 and Z. In general, the event of Drainage affects the node Saturated Thickness.

Finally, it is evidence that slope sliding along different slip surfaces and accordingly the failure surface may be of any shape [19, 57]. Undoubtedly, an existing form of slope failure can be defined as a result event in BNs. So, in this study, a translational failure event proposed is represented by a bottom node, named Slope Failure in Figure 6 , whose probability can be computed by a limit state function G(X),

$$G(X) = FOS - 1 \tag{10}$$

in which the node is a discrete variable with two states: $G(X) > 0$, the node denotes the probability of a stable slope, otherwise, it is the failure probability of the slope.

3.3 Slope failure analysis with Credal networks

In the analysis of an infinite slope with Bayesian networks approach, each node can be defined with precise probability distribution functions. With the limitation of information for the geotechnical problems, however, it is impossible to define all the nodes as point probabilities. Imprecise probabilities are employed to cope with imprecise problems in engineering analysis by means of transforming imprecise information into the probabilistic form [58]. This approach is also interesting to solve geotechnical problems. A discussion and comparison about how to define the input information using intervals or probabilistic model can be found in [59]. Interval model is the most common way to describe imprecision in geotechnical engineering, such as uncertain soil parameters with non-probabilistic information can be set as the form of the interval to make uncertainty analysis of a slope stability [60]. In this paper, hence, this approach is used to define the scarce input information. A Credal network proposed in this section is to estimate the probability of slope failure with limited information subject to drainage influence and initially demonstrate the suitability of CNs for the slope stability analysis.

As is aforementioned, the shear resistance cohesion and friction angle are two key factors of influencing the slope stability. If there is scarce information provided for an infinite slope, for example, the parameters c and ϕ change with geological/geotechnical conditions, so without any experiment test, it cannot be known in advance the exact properties of the slope's material. In this case, they should be two imprecise inputs for the CN.

The model implemented presents nine nodes, including discrete variables, continuous variables, interval variables and parametric p-boxes. These corresponding nodes are represented by rectangular, circle, ellipse and trapezoid, respectively (see Figure 7). In addition, the events and construction of the model remain similar to the previous BN model of the slope. Differently, the nodes Cohesion, Friction Angle, Saturated Unit Weight and Unsaturated Unit Weight associated with imprecise information, and the prior interval-value definitions of these imprecise nodes are achieved from expert judgement.

If the further information is available, such as the information of parametric p-boxes is linked with the nodes Cohesion and Friction Angle. To be specific, the two soil parameters can be described by the known probability distribution, whose parameters are interval variables. Then in this credal network, the imprecise information is presented by a combination of the nodes Vcohesion and Cohesion, Vfriction and Friction Angle. Comparing to the previous BN, the nodes Cohesion and Friction Angle in the CN model are substituted by the respective parametric p-boxes.

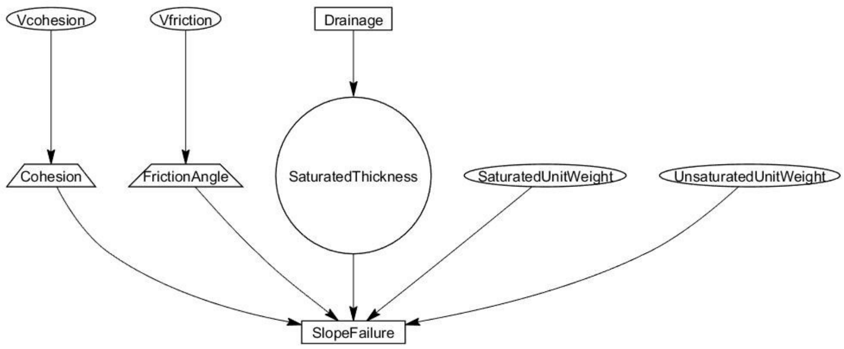

Figure 7: The CN model of an infinite slope.

Slope Failure is the node of interest in the CN, whose failure state of the node can predict the occurrence of a shallow landslide. The probability of slope failure is inferred by marginal probability calculation in the reduced CN. Furthermore, an analysis can be conducted to demonstrate the effect of the node Drainage on the slope stability. A simple example is implemented in the next section in order to illustrate the feasibility of this method.

4 Example 1: soil slope

The translational slip shown in Figure 5 is adopted to illustrate example 1. The total thickness of the slope is 4 m at the inclination angle β=30°. The BN model in Figure 6 is adopted in this example.

In the view of quantification, the key parameters of the soil slope in the BN are all set as random variables with known probability distributions according to their uncertain properties. Amongst these, the thickness of unsaturated soil (Z_d) changes complimentarily with the thickness of the saturated slope,

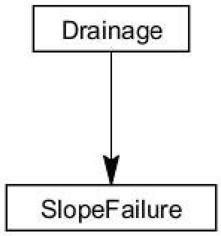

Figure 8: Reduction section of the network modeling soil slope.

and thus can be expressed by $(4 - Z_{sat})$. The detailed definitions of variables involved in the BN can be obtained from Table 1. The CPT of querying node of slope failure is evaluated by means of reliability analysis, which is used to update the model with inserting evidence as well.

In light of the above-mentioned method, the CPT of the querying node can be computed by equation (11) to (12), which is shown as follows,

$$P(SF) = P(D) \int \int_{\Omega_{SF}(\gamma_d, \gamma_{sat}, Z_{sat}, c, \phi)} f(\gamma_d) f(\phi) f(c) f(\gamma_{sat}) f(Z_{sat}|D)\, d\gamma_d\, d\phi\, dc\, d\gamma_{sat}\, dZ_{sat} \tag{11}$$

here, the domain of the slope failure $\Omega_{SF}(\gamma_d, \gamma_{sat}, Z_{sat}, c, \phi)$ can be described by the limited state function according to equation(10)), as:

$$P(SF) = \begin{cases} P_f, & FOS - 1 < 0 \\ P_s, & FOS - 1 \geq 0 \end{cases} \tag{12}$$

in which P_f denotes the failure probability of the slope while the safe probability is P_s.

Table 1: Input parameters of the unsaturated soil.

Parameters	Variable type	CPD*
Cohesion [kPa]	Continuous	logN(22,10)
Friction angle [°]	Continuous	N(35,3)
Unsaturated unit weight [kN/m^3]	Continuous	N(17, 0.4)
Saturated unit weight [kN/m^3]	Continuous	N(19, 0.5)
Saturated thickness of soil [m]	Continuous	U(0, 4) or 0
Drainage (D)	Discrete	[0.5, 0.5]
Slope failure(SF)	Discrete	[P_f,P_s]

*N, logN, represents normal and lognormal distribution with mean and standard deviation, respectively. U represent uniform distribution with lower and upper bound.

In order to characterize the relationship between slope stability and other influence factors, the easy way is to change these nodes to obtain a new result of the slope stability. Then the new observations with narrowing the distribution range of random variables are inserted into the continuous nodes. On the basis of the expert knowledge, the new bounds of the nodes Cohesion, Friction Angle, Unsaturated Unit Weight, and Saturated Unit Weight, are [0, 100], [25, 45], [16, 19], [18, 21], respectively. Besides, a narrower bound of each observed node also is computed to make a further observation. Monte Carlo simulation is applied to each network updating.

4.1 Results from example 1

Figure 8 shows the reduced BN, where all the continuous variables are removed from the original BN by means of eBNs approach, and only two discrete nodes left: Drainage and Slope Failure. Then the reasoning in the eBN can be inferred with this traditional BN, and the results are exhibited in Table 2. The prior marginal probability of node Slope Failure is 2.74%. Compared with no evidence, the occurrence of failure of the slope in case of drainage is rare, which is much lower than the state of no drainage, whose result with 5.13%. That means that if drainage takes place it can stabilize the slope. Furthermore, it can be reasonably achieved that drainage is of importance to this slope. Then from this information support, the decision maker immediately knows the risk of slope failure can be avoided if he spends money in draining the slope.

Table 2: The effect of Drainage on slope safety.

State	No evidence	Drainage	No Drainage
$P(FS)$	0.0274	$< 10^{-7}$	0.0513

In geotechnical problems, it is common that the soil characterization is performed in different phases and, therefore, new observations can be obtained in an advanced step of the study. These new results serve to identify the influence of soil parameters on the slope stability. The adoption of the discretized approach allows considering these new results as evidence, updating the probabilities in the model. From the results in Table 3, the failure probability of the slope varies from 2.55% to 2.67%, which is close to the original result, but the new information indicates the reduced effect on the failure result.

Table 3: Slope failure probability updated with new information.

Nodes	Cohesion	Friction Angle	Unsaturated Unit Weight	Saturated Unit Weight
Evidence	$0 \le c \le 100$	$25 \le \phi \le 45$	$16 \le \gamma_d \le 19$	$18 \le \gamma_{sat} \le 21$
$P(FS)$	0.0255	0.0267	0.0259	0.0256

Table 4: Slope failure probability updated with further information.

Nodes	Cohesion	Friction Angle	Unsaturated Unit Weight	Saturated Unit Weight
Evidence	$10 \le c \le 90$	$30 \le \phi \le 40$	$16.5 \le \gamma_d \le 18.5$	$18.5 \le \gamma_{sat} \le 20.5$
$P(FS)$	0.0181	0.0237	0.0262	0.0253

The reason for such a small variation in the failure probability results from the large range given for the observations. Therefore, the outcome will be much more evidence if the observations are narrower, what could be a result of additional geotechnical tests, for example, narrowing the considered intervals as Table 4 shows. $P(FS)$ is 1.81% and 2.37% respectively with the corresponding limited range of the nodes Cohesion [10, 90] and Friction Angle [30, 40]. These results reveal that both of them greatly affect the reduction of the risk of the slope, while Unsaturated Unit Weight and Saturated Unit Weight have a smaller effect on the results of $P(FS)$ in comparison with the results of Table 3. Given a slight change in each observed nodes, the value of $P(FS)$ shows more apparently variation in Cohesion and Friction Angle than nodes Unsaturated Unit Weight and Saturated Unit Weight. In this study, therefore, Cohesion and Friction Angle are more important for this slope stability.

5 Example 2: slope in residual soil from Porto

In regions where igneous rock, like granite or gneiss are present, the weathering of the rock produces soils designated as residual soils. These materials are very common in mountainous countries as the case of Portugal, Spain, Brazil, China, Hong Kong, Singapore, and Africa.

In the northern part of Portugal, there has been an extensive geotechnical characterization of granite residual soils [61–63]. Table 5 presents the common strength parameters found in the residual soil from Granite, in the Porto region. The typical values for strength parameters of this type of soil from Porto, such as cohesion and friction angle, are known as the interval-value with lower and upper bounds. Unsaturated and saturated unit soil weight are both defined based on expert knowledge. Besides, for a typical design, a slope in residual soils is typically designed with a fixed inclination of 3H to 2V, and the total thickness of this slope is assumed as 4 m in this study. The failure surface is considered parallel to the surface of the slope, as shown in Figure 9.

Three different situations of information available in Cohesion, Friction Angle, Unsaturated Unit Weight, and Saturated Unit Weight are described in Table 5. With the model, a change in any of the four factors would cause the varying failure probability of a slope. For precise information provided, the eBNs approach is used here as a reference, and interval analysis is adopted to cope with the limited information. If further information about the variables can be achieved, such as input distribution with a bound on its mean, then the parametric p-boxes might be introduced in the imprecise nodes Cohesion and Friction Angle. Thus it is possible to observe the change of the results in comparison with only interval nodes in the model.

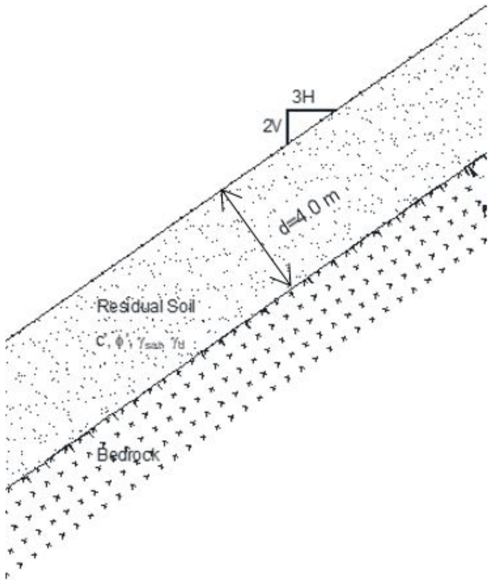

Figure 9: A residual soil slope.

Table 5: Input parameters of the residual soil.

Parameters	Precise information	Imprecise information min	Imprecise information max	P-boxes*
Cohesion [kPa]	logN(20,4)	0	70	logN(μ_c,4), $\mu_c \in [16, 22]$
Friction angle [°]	N(37,1.85)	25	47	N(μ_f,1.85), $\mu_f \in [36, 38.5]$
Unsaturated unit weight [kN/m^3]	N(18.5,0.51)	17	20	[17, 20]
Saturated unit weight [kN/m^3]	N(20,0.6)	18	22	[18, 22]

$*\mu$ indicates the mean of the distributions. The notes of Table 1 also apply here.

Considering the Credal network of the residual soil slope, the CPT of the discrete node Slope Failure is obtained on the basis of the factor of safety, same as equation (11) and (12), where the calculation of the failure probability is made with adaptive line sampling approach [49].

5.1 Results from example II

From the results (Table 6), it can be seen that an exact probability of slope failure can be obtained with the precise input for the conservative model. The 2.47% failure probability indicates a reasonable risk for this existing slope with precisely specific parameters. However, in the case of poor information, the input uncertainty affects the precision of output so that the results are denoted with the probability bounds. When the input nodes Cohesion, Friction Angle, Unsaturated Unit Weight, and Saturated Unit Weight only can be defined as interval variables with the limited information, the probability bound of slope failure is between 0 and 1. The result is too wide to provide the useful information of slope stability. In other words, each combination of the different values of the factors can produce any possibility of the slope states, failure or safe. Hence, the feasible way is to reduce the uncertainty input in order to increase the precision of the output, what can be done by producing additional geotechnical information or by approaching the reliability problem with different methods. For example, a practical common geotechnical solution would result from performing additional boreholes in the slope and laboratory test what would allow to more precise geotechnical parameters.

Table 6: Slope failure probability.

Different information	Precise	Imprecision	Parametric p-box
$P(FS)$	0.0274	[0, 1]	[0, 0.0711]

Comparing to the first two input information, the further observation is to add the probability boxes in the imprecise nodes Cohesion and Friction Angle. As it is shown in Table 6, the probability bound of failure slope became dramatically tighter after introducing P-boxes. The range of the failure result is from 0 to 7.11%, and the upper bound of the failure slope reveals a steep decrease. Besides, the precise result with 2.47% is included in this range. It illustrates actual risk can be estimated with the consideration of the reasonable application of parametric p-boxes in the CN model.

Table 7: Failure probabilities with two states of Drainage.

Different information	Precise	Imprecision	Parametric p-box
$P(FS\|D = true)$	0	[0, 1]	0
$P(FS\|D = false)$	0.0514	[0, 1]	[0, 0.0153]

In Table 7, the risk of slope failure under drained conditions shows a greatly reduced tendency, and even the risk can decrease to 0 in contrast to the state of no drainage. That is because if water is away, the percolation forces disappear and the resistant forces also increase, as a result of the increase in the normal force and, therefore, the friction component of the strength also increases.

The result with the interval [0, 1] based on the very poor information cannot give further information for decision makers, but the probability bound of Slope Failure with the evidence Drainage makes sense by ways of p-boxes. Specifically, if drainage is not implemented, the failure result of the residual soil slope with [0, 15.31%] is much wider than the one with drainage.

6 Conclusions

This study presents the advanced Bayesian networks methods to estimate the risk of failure of the slope subjected to drainage state. For the purpose of identifying the factors that affect the slope failure the most, new observations are made in some continuous nodes to update the model. The proposed methods proved useful and the results provide significant information for the decision makers.

Enhanced Bayesian networks and Credal networks are applied relying on the input information availability. Enhanced Bayesian networks consist of two types of nodes, continuous and discrete nodes, where an integration of Bayesian networks and structural reliability analysis is applied to make the inference in this precise model, while the Credal network, especially for the scenario that there is no enough abundant information to get the precise CPDs for each of nodes. Additionally, discrete variables, random variables, interval variables and p-boxes are presented in the model. After the adoption of adaptive line sampling, the reduced model is computed by exact inference algorithms. The bounds of result provide a rough estimation of the actual risk, and the permission of application of p-boxes in the model contributes to the reduction of the uncertainty in output. Moreover, a discretization process is applied when new evidence enters the continuous nodes. These capabilities ensure the wide flexibility of the model in analyzing the risk of the slope.

The two examples implemented demonstrate that the models have interesting possibilities for assessing the risk of the slope. The exact failure probability of an infinite slope in the first example indicates a low risk, and according to the analysis of updating the information in the specific nodes, the conclusion can be made that the risk of the slope can be significantly improved with drainage. Although interval-value is a suitable way for representing the non-probabilistic information, the interval results of the residual soil slope may fail to acquire the usable range of real value. In this case, p-boxes involved obviously narrow the bound of failure probability. All in all, both of enhanced Bayesian networks and Credal networks are effective and feasible means to estimate the failure of a slope.

Acknowledgments

This work is supported by Chinese Scholarship Council.

References

[1] H. Rahardjo, S. N. Alfrendo, C. L. Eng, and Y. S. Ng, Effects of groundwater table position and soil properties on stability of slope during rainfall, *Geotechnical and Geoenvironmental Engineering* **136**, 1555 (2010).

[2] J. Ma, H. Tang, X. Hu, A. Bobet, M. Zhang, T. Zhu, Y. Song, and A. M. Mutasim, Identification of causal factors for the Majiagou landslide using modern data mining methods, *the International Consortium on Landsli* **14**, 311 (2016).

[3] N. Lu and J. Godt, Infinite slope stability under steady unsaturated seepage conditions, *Water Resources Research* **44**, 404 (2008).

[4] R. L. Ray, J. M. Jacobs, and P. de Alba, Impacts of unsaturated zone soil moisture and groundwater table on slope instability, *Journal of Geotechnical and Geoenvironmental Engineering* **136**, 1448 (2010).

[5] A. Zhou, J. Huang, and C. Q. Li, Failure analysis of an infinite unsaturated soil slope, *Geotechnical Engineering* **169**, 410 (2016).

[6] Y. Oka and T. H. Wu, System reliability of slope stability, *Geotechnical Engineering* **116**, 1185 (1990).

[7] P. J. Finlay, R. F. and, and P. K. Maguire, The relationship between the probability of landslide occurrence and rainfall, *Canadian Geotechnical Journal* **34**, 811 (1997).

[8] H. El-Ramly, N. R. Morgenstern, and D. M. Cruden, Probabilistic slope stability analysis for practice, *Canadian Geotechnical Journal* **39**, 665 (2002).

[9] J. Zhang, H. W. Huang, L. M. Zhang, H. H. Zhu, and B. Shi, Probabilistic prediction of rainfall-induced slope failure using a mechanics-based model, *Engineering Geology* **168**, 129 (2013).

[10] D. Zhu, D. V. Griffiths, J. Huang, and G. A. Fenton, Probabilistic stability analyses of undrained slopes with linearly increasing mean strength, *Geotechnique* **67**, 733 (2017).

[11] S. Metya, T. Mukhopadhyay, S. Adhikari, and G. Bhattacharya, System reliability analysis of soil slopes with general slip surfaces using multivariate adaptive regression splines, *Computers and Geotechnics* **87**, 212 (2017).

[12] S. Metya, T. Mukhopadhyay, S. Adhikari, and G. Bhattacharya, Efficient evaluation of reliability for slopes with circular slip surfaces using importance sampling, *Geotechnical and Geoenvironmental Engineering* **135**, 768 (2009).

[13] Y. Wang, Z. J. Cao, and S. K. Au, Practical reliability analysis of slope stability by advanced monte carlo simulations in a spreadsheet, *Geotechnical and Geoenvironmental Engineering* **48**, 162 (2011).

[14] M. G. Sakellariou and M. D. Ferentinou, A study of slope stability prediction using neural networks, *Geotechnical and Geological Engineering* **23**, 419 (2005).

[15] A. Chakraborty and D. Goswami, Slope stability prediction of slope stability using multiple linear regression (MLR) and artificial neural network, *Arabian Journal of Geosciences* **10**, 385 (2017).

[16] W. J. Liang, D. F. Zhuang, and D. Jiang, Assessment of debris flow hazards using a bayesian network., *Geomorphology* **171**, 94 (2012).

[17] Y. Q. Song, J. H. Gong, and S. Gao, Susceptibility assessment of earthquake-induced landslides using bayesian network: A case study in Beichuan, China, *Computers and Geosciences* **42**, 189 (2012).

[18] T. J. Jiang and D. L. Wang, A landslide stability calculation method based on bayesian networks, *2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA)* (2013).

[19] W. S. Liu, S. B. Li, and R. Tang, Slope stability evaluation in open pit based on bayesian networks, *Proceedings of the ICCEAE2012* **181**, 1227 (2013).

[20] M. Peng, X. Y. Li, and D. Q. Li, Slope safety evaluation by integrating multi-source monitoring information, *Structural Safety* **49**, 65 (2013).

[21] J. Pearl, *Probabilistic reasoning in intelligent systems* (Morgan Kaufmann Publishers., San Mateo, California, 1988).

[22] D. Straub, Natural hazards risk assessment using bayesian networks, in *Proceedings of the 9th International Conference on Structural Safety and Reliability(ICOSSAR)* (2005).

[23] J. Pearl and S. Russell, *Causality: models, reasoning, and inference* (Cambridge University Press, 2000).

[24] D. Straub and A. Der Kiureghian, Combining bayesian networks with structural reliability methods: Methodology, *Journal of Engineering Mechanics* **136**, 1248 (2010).

[25] A. Antonucci, A. Salvetti, and M. Zaffalon, Assessing debris flow hazard by credal nets, *2nd International Conference on Soft Methods in Probability and Statistics* 125–132 (2004).

[26] F. Jensen, *An introduction to Bayesian networks* (Springer-Verlag, Heidelberg Press, 1996).

[27] P. Weber, G. Medina-Oliva, C. Simon, and B. Iung, Overview on bayesian networks applications for dependability, risk analysis and maintenance areas, *Engineering Applications of Artificial Intelligence* **25**, 671 (2012).

[28] R. Daly, Q. Shen, and Q. S. Aitken, Learning bayesian networks: approaches and issues, *The Knowledge Engineering Review* **26**, 99 (2011).

[29] F. Jensen, *Bayesian networks and decision graphs* (Springer-Verlag, Heidelberg Press, 2001).

[30] F. Jensen, *Modeling and reasoning with Bayesian networks* (Cambridge University Press, 2009).

[31] H. Langseth, T. D. Nielsen, R. Rumi, and A. Salmerón, Inference in hybrid bayesian networks, *Reliability Engineering System Safety* **94**, 1499 (2009).

[32] D. Straub and A. Der Kiureghian, Combining bayesian networks with structural reliability methods: Application, *Journal of Engineering Mechanics* **136**, 1259 (2010).

[33] S. Tolo, E. Patelli, M. Beer, and M. Broggi, Enhanced bayesian networks approach to risk assessment of spent fuel ponds, in *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering(ICASP12)* (2015).

[34] J. Dougherty, R. Kohavi, and M. Sahami, Supervised and unsupervised discretization of continuous features, *In: Proceedings of the 12th International Conference on Machine Learning (ICML)* (1995).

[35] L. Kurgan and K. Cios, Discretization algorithm that uses class-attribute interdependence maximization, in *Proceedings of the International Conference on Artificial Intelligence (IC-AI)* (Las Vegas, Nevada, 2001), 980–987.

[36] K. Zwirglmaier and D. Straub, A discretization procedure for rare events in bayesian networks, *Reliability Engineering and System Safety* **152**, 96 (2016).

[37] Y. C. Chen, T. A. Wheeler, and M. J. Kochenderfer, Learning discrete bayesian networks from continuous data, *Journal of Artificial Intelligence Research* **59**, 103 (2017).

[38] E. Fagiuoli and M. Zaffalon, 2U: An exact interval propagation algorithm for polytrees with binary variables, *Artificial Intelligence* **106**, 77 (1988).

[39] F. G. Cozman, Credal networks, *Artificial Intelligence* **120**, 199 (2000).

[40] A. Antonucci, C. P. de Campos, and M. Zaffalon, in *Introduction to Imprecise Probabilities*, edited by T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes (John Wiley & Sons, 2014).

[41] B. Tessem, Interval probability propagation, *International Journal of Approximate Reasoning* **7**, 95 (1992).

[42] J. C. Ferreira da Rocha and F. G. Cozman, Inference with separately specified sets of probabilities in credal networks, in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence* (San Francisco, 2002), 430–437.

[43] J. S. Ide and F. G. Cozman, Approximate algorithms for credal networks with binary variables, *International Journal of Approximate Reasoning* **48**, 275 (2007).

[44] W. Y. Liu and K. Yue, Bayesian network with interval probability parameters, *International Journal on Articial Intelligence Tools* **20**, 911 (2011).

[45] A. Antonucci, C. de Campos, D. Huber, and M. Zaffalon, Approximate credal network updating by linear programming with applications to decision making, *International Journal of Approximate Reasoning* **58**, 25 (2015).

[46] S. Tolo, E. Patelli, and M. Beer, An open toolbox for the reduction, inference computation and sensitivity analysis of credal networks, *Advances in Engineering Software* **115**, 126 (2018).

[47] S. Tolo, E. Patelli, and M. Beer, An inference method for bayesian networks with probability intervals, in *Proceedings of the joint ICVRAM-ISUMA UNCERTAINTIES Conference* (Florianópolis, SC, Brazil, 2008).

[48] Y. Xia, M. Mahmoodian, C. Q. Li, and A. Zhou, A stochastic method for predicting risk of slope failure subjected to unsaturated infiltration flow, *International Journal of Geomechanics* **17** (2017).

[49] M. De Angelis, E. Patelli, and M. Beer, Advanced line sampling for efficient robust reliability analysis, *Structural Safety* **52**, 170 (2014).

[50] P. S. Koutsourelakis, Reliability of structures in high dimensions.part ii. theoretical validation, *Probability Engineering Mechnique* **19**, 19 (2004).

[51] A. W. Bishop, The use of the slip circle in the stability analysis of slopes, *Geotechnique* **5**, 7 (1955).

[52] N. R. Morgenstern and V. Price, The analysis of the stability of general slip surface, *Geotechnique* **15**, 79 (1965).

[53] S. Xiao, L. Yan, and Z. Cheng, A method combined numerical analysis and limit equilibrium theory to determine potential slip surfaces in soil slopes, *Journal of Mountain Science* **8**, 718 (2011).

[54] W. Niu, Determination of slope safety factor with analytical solution and searching critical slip surface with genetic-traversal random method, *Scientific World Journal* (2014).

[55] G. Acharya, F. De Smedt, and N. T. Long, Assessing landslide hazard in GIS: A case study from Rasuwa, Nepal, *Bulletin of Engineering Geology and the Environment* **65**, 99 (2006).

[56] K. Terzaghi, *Theoretical soil mechanics* (John Wiley & Sons, New York, 1943).

[57] D. Q. Li, Z. Y. Yang, Z. J. Cao, S. K. Au, and K. K. Phoon, System reliability analysis of slope stability using generalized subset simulation, *Applied Mathematical Modelling* **46**, 650 (2017).

[58] M. Beer, S. Ferson, and V. Kreinovich, Imprecise probabilities in engineering analyses, *Mechanical Systems Signal Processing* **37**, 4 (2013).

[59] M. Beer, Y. Zhang, S. T. Quek, and K. K. Phoon, Reliability analysis with scarce information: comparing alternative approaches in a geotechnical engineering context, *Structural Safety* **41**, 1 (2012).

[60] E. Rubio, J. W. Hall, and M. G. Anderson, Uncertainty analysis in a slope hydrology and stability model using probabilistic and imprecise information, *Computers and Geotechnics* **31**, 529 (2004).

[61] A. Viana da Fonseca and R. Coutinho, Characterization of residual soils, in *Proceedings of the 3rd International Conference on Site Characterization,* (2008), 195–248.

[62] A. Topa Gomes, Elliptical shafts by the sequential excavation method. the case of Metro do Porto, (in portuguese), *PhD Thesis, University of Porto* 195–248 (2009).

[63] L. Pinheiro Branco, A. Topa Gomes, A. Silva Cardoso, and C. Santos Pereira, Natural variability of shear strength in a granite residual soil from Porto, *Geotechnical and Geological Engineering* **32**, 911 (2014).