Uncertainty Tracking and Geotechnical Reliability Updating Using Bayesian Networks

Duvvuru Mohan, Varenya; Vardon, Phil; Hicks, Michael; van Gelder, Pieter

DOI
10.3850/978-981-11-2725-0-IS3-5-cd

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019)

Citation (APA)
Duvvuru Mohan, V., Vardon, P., Hicks, M., & van Gelder, P. (2019). Uncertainty Tracking and Geotechnical Reliability Updating Using Bayesian Networks. In J. Ching, D.-Q. Li, & J. Zhang (Eds.), *Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019): State-of-the-Practice in Geotechnical Safety and Risk* (pp. 631-636). Article IS3-5. https://doi.org/10.3850/978-981-11-2725-0-IS3-5-cd

Uncertainty Tracking and Geotechnical Reliability Updating Using Bayesian Networks

Varenya Kumar D. Mohan[1], Philip J. Vardon[1], Michael A. Hicks[1], and
Pieter H.A.J.M. van Gelder[2]

[1]Geo-Engineering Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, the Netherlands.
E-mail: v.k.duvvurumohan@tudelft.nl , p.j.vardon@tudelft.nl , m.a.hicks@tudelft.nl
[2]Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands.
E-mail: P.H.A.J.M.vanGelder@tudelft.nl

Abstract: Bayesian networks are proposed as a tool to integrate reliability and influential variables relating to the slope stability of an idealized embankment. The site investigation (extent) and slope geometry, as well as the material properties and their spatial variability, are considered within a Bayesian network. The random finite element method (RFEM) is used to quantify the slope reliability and demonstrate the overall methodology. Prior probabilities of geometry, material parameters and their heterogeneity are obtained from 'initial' site investigation data. Probabilistic distributions of slope performance (factor of safety) are obtained by Bayesian inference in the network to investigate the impact of additional site investigation. The amount of additional site investigation required to increase the geotechnical reliability is assessed. This work illustrates the applicability of Bayesian networks as an effective reliability and uncertainty assessment tool that can aid decision making for site investigation and during maintenance, where new observations can be readily integrated to obtain updated reliability estimates.

Keywords: Bayesian network; geotechnical reliability; random fields; slope reliability; spatial variability; uncertainty.

1 Introduction

Data from site investigation can be used to characterize the spatial variability of soil parameters. Statistical distributions of various parameters can be incorporated into geotechnical reliability modeling using methods such as the random finite element method (RFEM). RFEM uses Monte Carlo analyses, with random fields to model the spatial variability of soil properties and the finite element method to analyze geotechnical performance (Hicks and Samy 2004; Fenton and Griffiths 2008). Significant advances have been made in improving the methodology and reducing geotechnical uncertainty (e.g., Li et al. 2016; Vardon et al. 2016). While RFEM is an effective procedure to incorporate spatial variation in properties, it remains computationally expensive. Therefore, it is not always economically feasible to update geotechnical reliability estimates whenever new site investigation data become available. Furthermore, when additional site investigation is planned, it is difficult to assess the reduction of uncertainty that will be achieved from such efforts. In other words, without intensive computations, one cannot predict the extent of site investigation required for achieving a desired increase in reliability. The aforementioned two problems can be solved by developing a less expensive surrogate model that can represent slope reliability, or at least, approximately reproduce the dependencies between parameters. This paper proposes the use of Bayesian networks for developing a site-specific surrogate model that addresses these two needs – (i) updating reliability estimates when new information is made available; (ii) evaluating the extent of additional site investigation that is required and the value it provides. In particular, the use of Bayesian networks that allow for input of continuous probability distributions are investigated, which enable: (i) representation of thorough reliability methods (e.g., RFEM) where continuous probability distributions act as inputs; and (ii) uncertainty quantification and tracking through statistical measures (e.g., coefficient of variation or standard error).

In civil engineering, Bayesian networks have been extensively used for reliability estimation and updating in structural engineering applications (e.g., Straub and Kiureghian 2010; Luque and Straub 2016). Applications in geotechnical engineering have been less frequent, but nevertheless attempted, for estimating geotechnical reliability and risk. Nadim (2017) presented the developments and challenges in the use of Bayesian networks for reliability-based geotechnical design. Chivatá Cárdenas (2019) proposed statically discretized Bayesian networks as a meta-modeling approach for analyzing slope stability, in which finite element modeling data were used to train the network as a surrogate model. Delgado-Hernandez et al. (2012) and Morales-Napoles et al. (2014) used Non-Parametric Bayesian Networks (NPBNs) to assess risks to dams in Mexico. The NPBN approach allows for modelling of continuous random variables in the Bayesian network without the arduous (and potentially error-prone) population of conditional probability tables. However, it is often limited to the use of the normal copula for defining dependencies between variables (Hanea et al. 2015). Peng et al. (2014) proposed a Bayesian network-based method for the evaluation of slope safety when monitoring data are available from

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_IS3-5-cd

several sources. They also recommended further consideration of the impact of spatial variability of soil properties and correlation of measurements on slope safety. Zhang et al. (2015) applied a similar methodology for practical cases of slope stability. Li et al. (2018) examined slope stability with large amounts of monitoring data, and used Bayesian networks to model correlated material properties and observations. Further consideration of spatial variability of material properties was recommended by the authors. Jiang et al. (2018) presented a structural reliability-based method to account for spatial variability in Bayesian updating, integrated with reliability estimation. This method, however, is not based on graphical models such as the Bayesian network. Roscoe and Hanea (2015) and Roscoe (2017) modelled the reliability of a system of levees using NPBNs, focusing on the piping mechanism of failure. They included an aspect of spatial variability (in material properties) that affects the reliability of the levee system, by considering variability across homogenous sections. Data of strong performance of the levees under high water levels was included in the network, resulting in a significant improvement in the posterior estimates of reliability. Thus, Bayesian networks have been shown to be particularly useful in the context of reliability estimation for: (i) representing a multivariate problem with correlated variables in an efficient framework; (ii) modelling uncertain geotechnical information under data scarcity and tracking uncertainty; and (iii) updating reliability estimates when new input information is available.

2 Bayesian Networks – A Brief Introduction

A Bayesian network is a specific application of Bayesian probability theory. It is a directed acyclic graph, composed of 'nodes' that correspond to random variables and 'arcs' that link dependent variables. The directions of the arcs indicate the dependencies between the nodes (i.e. directed), and these arcs never cycle back from the child nodes to the parent nodes (i.e. acyclic). Hence, the network is a visually explicit representation of the mutual relationship between random variables, and represents the joint probability distribution (JPD) of all random variables within the model. A simple example is shown in Figure 1.

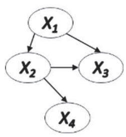

Figure 1. Example of a Bayesian network; X_i indicates a random variable.

The dependencies between random variables are usually encapsulated within conditional probability distributions (given by ($P(X_i|\text{Parents}(X_i))$) at each node. The JPD is given by the chain rule of Bayesian networks (see Eq.1) and the JPD of the example in Figure 1 is given by Eq. 2.

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P(X_i \,|\text{Parents}(X_i)) \tag{1}$$

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|\, X_1)P(X_3|X_1, X_2)P(X_4|\, X_2) \tag{2}$$

The JPD can be queried to infer the state of a random variable, given our beliefs regarding the other variables, via Bayesian inference. In other words, Bayesian networks can be used to answer probabilistic queries in a multivariate problem when one or more variables have been observed.

3 Site-Specific Bayesian Network Using RFEM Analysis

The 3D RFEM implementation from Varkey et al. (2019) is used as the basis for building the Bayesian network in this study. Accordingly, the slope used in that study is shown in Figure 2, and the results of the study are directly adopted as the reliability estimate from RFEM that is to be represented within the network.

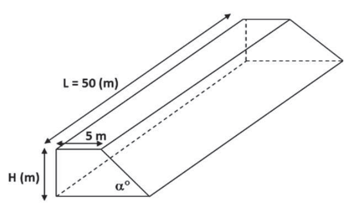

Figure 2. Model slope used in the 3D RFEM analysis in Varkey et al. (2019)

3.1 Selection of random variables within the Bayesian network

In Varkey et al. (2019), a range of slope heights and slope angles were considered, as shown in Table 1. The soil unit weight was 20 kN/m^3, Poisson's ratio 0.3 and Young's Modulus 100 MPa. Hence, these were deterministic parameters that are not considered as random variables within the network. The cohesion and friction angle of the soil were spatially variable and had normal distributions with means of 10 kPa and 25°, respectively, and a coefficient of variation of 0.2. From these original distributions, random fields of property values were generated for each RFEM realization. The standard deviations of strength properties within each realization were observed to have little correlation with the factor of safety. Hence, only the mean values of cohesion and friction angle are considered for each realization. The vertical scale of fluctuation was assumed to be constant (θ_v = 1 m), and hence is not considered as a node in the Bayesian network. A range of values were considered for the horizontal scale of fluctuation (θ_h) as shown in Table 1. These soil properties are assumed to be obtained from an 'initial site investigation' or a best estimate prior to the RFEM estimation of reliability. The factor of safety and failure length were obtained for each realization of the RFEM analysis. In all, the dataset used in this study is composed of 9000 realizations based on the different discrete combinations of slope height, slope angle and horizontal scale of fluctuation shown in Table 1.

The input random variables that need to be modelled in the Bayesian network are the same as in Varkey et al. (2019); namely, the slope height and angle, cohesion, friction angle and horizontal scale of fluctuation. The tangent of the friction angle is used instead of the friction angle, as it is more directly related to strength based on the Mohr-Coulomb failure criterion considered in this RFEM analysis. The output random variables from the RFEM analysis, i.e. factor of safety and failure length (as in Varkey et al. 2019), are also included in the network to complete the selection of random variables required to represent the RFEM analysis.

The impact of site investigation on reliability estimates is also considered; in this case, based on site investigation data obtained from cone penetration tests (CPTs). A decision variable "Number of CPTs" is introduced into the network as a quantitative measure of 'additional site investigation' that could be carried out for the slope. The variables considered as nodes in the network are listed in Table 1, along with their distribution type and parameters. Prior probabilities assumed at the nodes are discussed in Section 3.3.

Table 1. Inputs to 3D RFEM analysis in Varkey et al. (2019) and the Bayesian network in this study.

Random Variable	Inputs to RFEM analyses in Varkey et al. (2019)		Inputs to Bayesian Network in this study		
	Distribution Type	Parameters/Values	Discrete/ Continuous	Preferred Distribution Type	Parameters/ Functional Form
Cohesion (c)	Normal	$\mu = 10$, $\sigma = 2$ kPa	Continuous	Normal(1)	$\mu = 10$, $\sigma = 2$ kPa
Friction Angle (ϕ)	Normal	$\mu = 25$, $\sigma = 5°$	Continuous	Normal(1)	$\mu = 25$, $\sigma = 5°$
Horizontal Scale of Fluctuation (θ_h)	Discrete values	θ_h = 6, 12, 24 m	Continuous	Lognormal	$\mu = 1.15$, $\sigma = 0.87$ m
Slope Angle (α)	Discrete values	α = 26.6, 45.0, 63.4°	Continuous	Normal	$\mu = 45$, $\sigma = 1°$
Slope Height (H)	Discrete values	H = 3, 4, 5, 6 m	Continuous	Normal	$\mu = 5$, $\sigma = 0.1$ m
Factor of Safety	NA	NA	Continuous	Truncated Normal(2)	$\mu = 1.85-0.03\alpha-0.16*H+0.09c+1.39*tan\phi-0.0007\theta_h$; $\sigma = 0.09$; LB = 0, UB = 10 (3)
Failure Length	NA	NA	Continuous	Truncated Normal(2)	$\mu = -2.25-0.07\alpha+0.80*H+1.74c+2.87*tan\phi-0.13\theta_h$; $\sigma = 7.6$ m; LB = 0; UB = 50 m (3)
Number of CPTs (decision variable)	NA	NA	Definitional	-	N = 1 to 5(4)

1. Could be lognormally distributed, but normal assumption is suitable due to a low coefficient of variation (Table 1)
2. Distribution mean obtained based on EM learning algorithm in AgenaRisk®; linear Gaussian assumptions made for regression of distribution parameters (multivariate Gaussian assumptions not used for inference)
3. LB = Lower bound; UB = Upper bound for truncated normal distribution; assumed based on RFEM analysis results
4. Arbitrary assumption for demonstration of concept

3.2 Definition of Bayesian network structure

While the Bayesian network structure can be learnt from data (structure learning), there is sufficient understanding of the dependencies between random variables within the RFEM analysis. The structure of the Bayesian network is shown in Figure 3. A decision node, "Number of CPTs", is added later to demonstrate the impact of additional site investigation; it affects the uncertainty in the cohesion, tangent of the friction angle and horizontal scale of fluctuation, but there are no effects on the slope geometry variables.

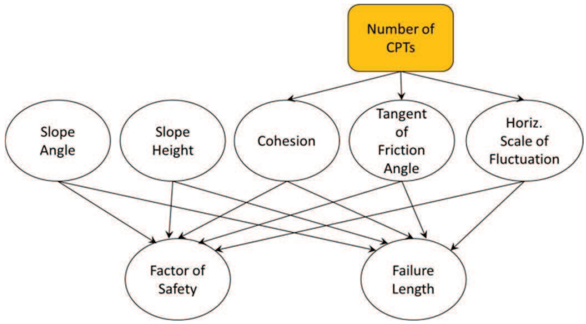

Figure 3. Bayesian network structure ("Number of CPTs" node added later as part of 'additional site investigation').

3.3 Definition of prior distributions at the nodes

The nodes in the network shown in Figure 3 (except the "Number of CPTs" node) are modelled as continuous variables to match the continuous distributions and inputs to the RFEM analysis. This also enables the tracking of the uncertainty via statistical measures of continuous probability distributions (e.g., standard deviation).

The RFEM analysis data from Varkey et al. (2019) are used for parameter learning to derive the conditional distributions (factor of safety and failure length). The Maximum Likelihood Estimation (MLE) via the Expectation Maximization (EM) algorithm is used here, as it is often available in commercial implementations. For continuous variables, this method assumes that all nodes in the network are normally distributed and the conditional relations are linear. Table 1 shows the functional form of the conditional linear Gaussian relations for the means of 'Factor of Safety' and 'Failure Length', as obtained from the EM algorithm implemented in AgenaRisk®. The conditional relations are assumed to be valid even if prior distributions are revised. While within the range of the initial distributions, this is a reasonable assumption, further investigation is not presented here, as the focus of this study is primarily to demonstrate the Bayesian network methodology. Parameter learning methods, free of the normality assumption, could be adopted.

The distribution types and parameters for the independent variables (marginal distributions) are chosen through judgement (see Table 1), and closely match the RFEM analysis. The horizontal scale of fluctuation is assumed to follow a lognormal distribution that includes the range of values used in the RFEM dataset. The distributions for slope height and angle are assumed to be normally distributed with parameters chosen to match the range of values in the RFEM dataset.

3.4 Dynamic discretization of prior distributions

As not all nodes are continuous and normally distributed, exact Bayesian inference is not possible. One common method to tackle this is to discretize the continuous variables. Static discretization, i.e., using a pre-defined, finite set of discrete states, may result in considerable loss of accuracy depending on the problem. The AgenaRisk® program uses a dynamic discretization algorithm that has greatly improved accuracy compared to static discretization (Fenton and Neil 2012). The dynamically discretized prior distributions are shown in Figure 4.

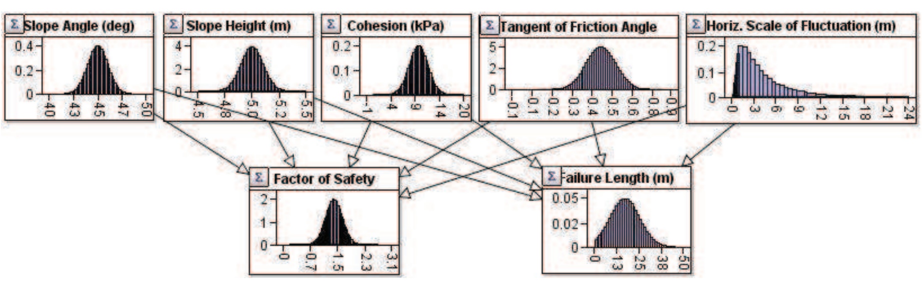

Figure 4. Dynamically discretized prior distributions.

3.5 Updating the network with 'additional site investigation' information

Additional site investigation is considered by adding the "Number of CPTs" node to the network. It is assumed that every additional CPT results in a reduction in standard deviation of the distributions of soil properties and the mean remains constant. While this is an assumption, in the case of existing slopes analyzed using RFEM, it is likely that a number of tests have already been performed and there is reasonable certainty in the mean value itself. For demonstration purposes, the % reduction in standard deviation ($\Delta\sigma$) is assumed to follow a simple linear relation for 1 to 5 additional CPTs:

$$-\Delta\sigma\ (\%) = (0.1\ \times \text{Number of CPTs} + 0.3)\ \times\ 100 \quad (3)$$

While this relation is arbitrarily assumed, it could also be modelled with analytical relationships from the literature relating number of datasets and statistical measures (e.g., for the COV of the scale of fluctuation (de Gast et al. 2018)). Given this dependence between the number of CPTs and the standard deviations of cohesion, friction angle and horizontal scale of fluctuation, a decision can be set at the "Number of CPTs" node to assess the impact on the distribution of factor of safety (measure of reliability/probability of failure) and the failure length (consequence measure). The junction tree algorithm was used for performing inference in the network.

4 Results and Discussion

Figure 5a shows the distribution of factors of safety obtained from the Bayesian network, for additional CPTs along with the prior distribution (zero additional CPTs). A progressive reduction in the standard deviation is observed with each additional test, as expected. Figure 5b shows the values of factors of safety with a 95% probability of being exceeded, i.e. factor of safety corresponding to 95% reliability.

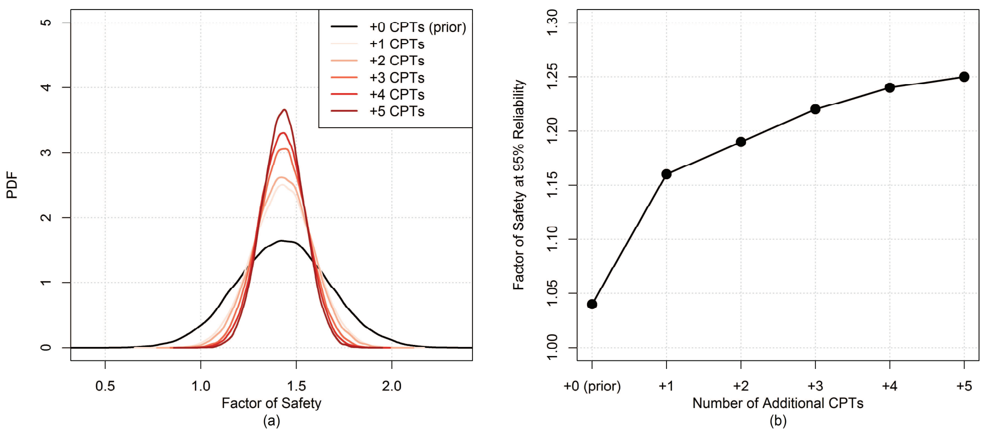

Figure 5. (a) Distribution of factor of safety; (b) Factor of safety at 95% reliability vs. no. of additional CPTs

It is seen that for zero additional CPTs (prior distributions), the factor of safety with a reliability of 95% is 1.04. By decreasing the uncertainty in soil properties through additional CPTs, say 4 tests, there is a greater than 18% increase in the factor of safety at 95% reliability. Hence, if the required factor of safety at 95% reliability is, e.g., 1.10, the slope would not pass an assessment before additional site investigation. However, with only 1 CPT the slope would be deemed stable. Thus, the Bayesian network allows for rapid sensitivity studies to estimate the extent of site investigation to achieve a minimum factor of safety at required reliability levels. Hence, it is also possible to directly evaluate the value of information from each additional CPT in financial terms. The confidence in the consequence of failure can be similarly estimated using the 'Failure Length' node of the Bayesian network. Diagnostic reasoning may also be performed by setting evidence for factor of safety or failure length and assessing the impact on individual slope input parameters.

Such use of the network is not necessarily limited to the number of additional tests. Nodes could be added to characterize the spatial distribution of testing locations which in turn affects the coefficient of variation of the soil property distributions (as in de Gast et al. 2018). Furthermore, the methodology described in this study can be used to develop similar surrogate models for generic 3D slope reliability analyses, if extensive datasets are available.

Three key areas in this study may be refined as part of future work. Firstly, the conditional probability relationships between input and output variables of the RFEM analysis were developed based on limited data from a previous study. The dataset could be expanded. Secondly, the parameter learning/regression methods for estimating the conditional probability distributions could be optimized for better fit with an expanded dataset. The prediction accuracy of the network can be assessed and increased this way. Finally, reduction in the uncertainty of input soil parameters due to additional CPTs (or other testing measurements) has been assumed for demonstrative purposes. During practical implementation of the methodology, the nature of the site and associated soil parameters need to be assessed.

5 Conclusions

Bayesian networks have been shown to be an effective methodology for representing slope reliability analysis, incorporating continuous distributions for soil properties and accounting for their spatial variability. Such a network can be used to plan site investigations, assess the value of additional observations and revise reliability estimates rapidly without the need for additional, computationally expensive and time-consuming numerical analyses.

Acknowledgments

This study has been carried out within the NARSIS project, which has received funding from the European Union's H2020-Euratom Program under grant agreement N° 755439.

References

Chivatá Cárdenas, I. (2019). On the use of Bayesian networks as a meta-modelling approach to analyse uncertainties in slope stability analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 13(1), 1-13.

de Gast, T., Vardon, P.J., and Hicks, M.A. (2018). Detection of soil variability using CPTs. *Cone Penetration Testing 2018*, Leiden: CRC Press, 289-294.

Delgado-Hernandez, D.J., Morales-Napoles, O., De-Leon-Escobedo, D., and Artega-Arcos, J.C. (2012). A continuous Bayesian network for earth dams' risk assessment: an application. *Structure and Infrastructure Engineering*, 10(2), 1-14.

Fenton, G.A. and Griffiths, D.V. (2008) *Risk Assessment in Geotechnical Engineering*, John Wiley & Sons, New York.

Fenton, N. and Neil, M. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks*, Boca Raton: CRC Press.

Hanea, A., Napoles, O.M., and Ababei, D. (2015). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety*, 144, 265-284.

Hicks, M.A. and Samy, K. (2004). Stochastic evaluation of heterogeneous slope stability. *Italian Geotechnical Journal*, 38(2), 54-66.

Jiang, S.H., Papaioannou, I., and Straub, D. (2018). Bayesian updating of slope reliability in spatially variable soils with in-situ measurements. *Engineering Geology*, 239, 310-320.

Li, Y.J., Hicks, M.A., and Vardon, P.J. (2016). Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields. *Computers and Geotechnics*, 79, 159-172.

Li, X., Zhang, L., and Zhang, S. (2018). Efficient Bayesian networks for slope safety evaluation with large quantity monitoring information. *Geoscience Frontiers*, 9(6), 1679-1687.

Luque, J. and Straub, D. (2016). Reliability analysis and updating of deteriorating systems with dynamic Bayesian networks. *Structural Safety*, 62, 34-46.

Morales-Nápoles, O., Delgado-Hernández, D.J., De-León-Escobedo, D., and Arteaga-Arcos, J.C. (2014). A continuous Bayesian network for earth dams' risk assessment: methodology and quantification. *Structure and Infrastructure Engineering*, 10(5), 589-603.

Nadim, F. (2017). Reliability-based approach for robust geotechnical design. *Unearth the Future, Connect beyond, 19th International Conference on Soil Mechanics and Geotechnical Engineering*, Seoul, 191-211.

Peng, M., Li, X.Y., Li, D.Q., Jiang, S.H., and Zhang, L.M. (2014). Slope safety evaluation by integrating multi-source monitoring information. *Structural Safety*, 49, 65-74.

Roscoe, K. and Hanea, A. (2015). Bayesian networks in levee reliability. *12th International Conference on Applications of Statistics and Probability in Civil Engineering*, Vancouver: cIRcle , 1-8.

Roscoe, K. (2017). *Bayesian Networks for Levee System Reliability: Reliability Updating and Model Verificatio*n. Delft: University of Technology.

Straub, D. and Der Kiureghian, A. (2010). Bayesian network enhanced with structural reliability methods: Application. *Journal of Engineering Mechanics*, 136(10), 1259-1270.

Vardon, P.J., Liu, K., and Hicks, M.A. (2016). Reduction of slope stability uncertainty based on hydraulic measurement via inverse analysis*. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(3), 223-240.

Varkey, D., Hicks, M.A., and Vardon, P.J. (2019). An improved semi-analytical method for 3D slope reliability assessments. *Computers and Geotechnics*, 111, 181-190.

Zhang, L., Li, X., Li, D., and Zhou, C. (2015). Interactive Evaluation of the Reliability of Engineered Slopes Utilising Multi-source Monitoring Information. *Geotechnical Safety and Risk V*, Amsterdam: IOS Press, 36-49.