Creative Commons Attribution

Periodica Polytechnica Civil Engineering

Optimizing Geotechnical Site Characterization: A Value of Information Approach Using Coupled Spatial Random Fields and Bayesian Networks

Sabah Hassan Fartosy[1*], Faris Waleed Jawad[1], Hayder A. Hasan[1]

[1] Department of Water Resources Engineering, Engineering College, Mustansiriyah University, 10052 Baghdad, P.O.B. 14022, Iraq
* Corresponding author, e-mail: dr.sabah77@uomustansiriyah.edu.iq

Received: 10 January 2026, Accepted: 18 May 2026, Published online: 02 June 2026

Abstract

Geotechnical design is greatly influenced by spatial variability of soil properties, uncertainty of modelling, and lack site investigation information. Traditional deterministic procedures do not directly define failure probability or the economic value of additional data, while strong design optimization may produce overly conservative solutions with no explicit cost–benefit evaluation. This study proposes an integrated probabilistic program that consolidates Spatial Random Fields (SRFs), Bayesian Networks (BNs), and Value of Information (VOI) analysis to support risk-informed geotechnical site characterization. To minimize the computational burden linked with high-dimensional random fields, the spatial variability is converted into minimized variables suitable for Bayesian inference and pre-posterior decision analysis. The proposed framework allows efficient determinations of the Expected Value of Sample Information (EVSI) and Expected Net Benefit of Sampling (ENBS) for alternative investigation protocols. Two representative case studies are studied: slope stability as an ultimate limit-state problem and shallow foundation settlement as a serviceability limit-state problem. The results reveal that the optimal sampling location is mechanism-dependent, shifting from the slope toe for stability assessment to the foundation center for settlement control. Validation against independent Monte Carlo simulations shows strong agreement, with ($R^2 \approx 0.95$) for the reduced order prediction. The proposed framework also produces up to (143%) improvement in economic efficiency compared with conventional investigation procedures. The framework therefore presents a practical-basis for economically optimized, risk-informed site investigation planning and future geotechnical digital twin applications.

Keywords

geotechnical risk, Bayesian Networks, spatial random fields, Value of Information, system reliability, slope stability

1 Introduction

The inherent heterogeneity of geologic materials introduces uncertainty into geotechnical engineering. The strength and stiffness of soil and rock formations can vary greatly on small spatial scales [1, 2]. The fact that traditional, deterministic design methods neglect spatial correlation of soil and rock properties can lead to designs that are either unnecessarily conservative or vulnerable to localized weak zones [3]. To account for spatial variability in geotechnical engineering, spatial random fields (SRFs) are typically used to model the ground [4, 5]. The work of Cho [6] and Santoso et al. [7] indicates that treating the entire soil layer as a single random variable ignores the spatial structure of the soil and underestimates the probability of failure. While creating a probabilistic model of the ground is the first and crucial step, the model must be continuously updated as data from investigations of the ground becomes available. The work of Papaioannou and Straub [8] indicates that using localized data from investigations can significantly reduce the risk of geotechnical failures. Based on this work, Bayesian Networks (BNs) became an interesting approach for modeling geotechnical problems. The work of He et al. [9] and Li et al. [10] specifically shows that BNs can be used as surrogate models for expensive simulations.

Despite the numerous developments in the field, there is a clear gap in the literature regarding the optimization of investigation plans. The Value of Information (VOI) analysis provides a solid theoretical basis for decision-making under uncertainty [11, 12]. The major challenge that VOI algorithms face is the high computational cost of pre-posterior

Cite this article as: Fartosy, S. H., Jawad, F. W., Hasan, H. A. "Optimizing Geotechnical Site Characterization: A Value of Information Approach Using Coupled Spatial Random Fields and Bayesian Networks", Periodica Polytechnica Civil Engineering, 2026. https://doi.org/10.3311/PPci.43308

analysis. In the context of SRF models, this makes pre-posterior analysis infeasible [13]. As a result, most pre-posterior site investigation optimization studies in the current literature are computationally forced to treat soil layers as single, non-spatial random variables [14]. Previous studies used BNs to carry out posterior analysis, but there is no efficient framework available for performing pre-posterior analysis for full VOI analysis. Such a framework is essential for engineering applications, such as determining the best location to drill a new borehole in a heterogeneous soil mass.

In this paper, a rigorous and validated methodology is proposed, integrating the spatial accuracy of SRFs and the computational efficiency of BNs. By using dimensionality reduction techniques, the high-dimensional SRF parameters are embedded into a BN. This enables conducting pre-posterior VOI analysis. Going beyond the concept of robust design [15], the developed and validated framework allows engineers to make informed decisions about the cost of investigations versus risk reduction. Overall, this framework provides a robust, decision-support tool for site investigations in soils with spatial variability.

2 Research background

Probabilistic Risk Assessment (PRA) combines the probability of failure and the consequence of failure mathematically [3]. PRA considers the uncertainty of parameters more suitably than deterministic values [16]. However, it differs from Robust Design Optimization (RDO), which attempts to identify a "worst case" scenario without quantifying probability, leading to overly conservative designs [15]. The proposed decision-theoretic framework is valuable not because it reduces the design based on RDO principles. Instead, it minimizes the Expected Total Cost by accounting for both the cost of failure and the cost of mitigation. A major shortcoming of state-of-the-art PRA treats site investigation as an input rather than a design variable. Spatial Random Fields (SRFs) mathematically describe the spatial continuity of soil property variability [4]. Ignoring the SRF structure leads to un-conservative designs, as the failure mechanism follows the correlated weak zones [5, 17]. Conditional Random Fields enhance the model by recognizing borehole data [18], but they are generated by an expensive Karhunen–Loève expansion and pose a serious constrain for iterative decisions.

Bayesian Networks (BNs) model variables and their conditional dependencies by using Directed Acyclic Graphs [19]. In geotechnical practice, BNs combine heterogeneous data to update risk assessments dynamically [20]. Standard BNs do instant inference and open exciting possibilities for the Digital Twin [21, 22]. However, they cannot handle continuous random fields because discretization leads to a surge of Conditional Probability Tables. Most work in geotechnics has simplified the problem by modeling soil strata as one variable, thus sacrificing spatial correlation to make efficient sampling designs feasible. Value of Information (VOI) analysis quantifies in economic terms the worth of collecting information before making a decision [11, 12]. It has not been applied to geotechnical engineering because the cost of pre-posterior analysis is prohibitive. It involves double simulation to assess the Value of Information by analyzing the worth of a measurement before it is made. For spatially variable soils, it is a computationally intractable problem [23, 24]. Yet current literature fails to connect high-fidelity Random Fields with efficient Bayesian inference. This research addresses this gap by proposing a framework that makes pre-posterior VOI analysis affordable even for complex geological contexts.

3 Problem definition and theoretical basis

Geotechnical systems are governed by uncertainty because of spatially varying soil parameters, uncertain loads, and the existence of multiple potential failure modes. Let $X(s)$ denote a vector field consisting of random variables governed over the domain Ω where $s \in \Omega$ represents space. Such random variables could be shear strengths, stiffnesses, and permeabilities. The spatial random field of $X(s)$ has associated hyperparameters θ that represent marginal means, variances, and correlation lengths which constitute a generalized probabilistic structure [4, 5]. Loads are represented by the variable L which could accrue to cumulative action (groundwater levels/rainfall/runoff/earthquake loading/excavation surcharge) characterized as uncertain loads/depths/rainfall intensities:

$$g_j\left(X(\cdot), L, d\right), \tag{1}$$

where d is a design/mitigation decision (geometry, reinforcement configuration, drainage configuration). Failure occurs in mode j if $g_j \leq 0$, where $F_j = \{g_j \leq 0\}$. For each failure mode, there is a consequence C_j, of which direct repair costs, indirect economic repercussions and, if applicable, social or environmental repercussions [3]. The cost of decision d is expressed by:

$$E\left[C_{\text{total}}(d)\right] = C_{cons}(d) + C_{inv} + P_f(d) \cdot C_{\text{fail}}, \tag{2}$$

where C_{cons} is the deterministic construction cost, C_{inv} is the cost of site investigation, $P_f(d)$ is the posterior probability of failure given the strategy, and C_{fail} is the monetized consequence of failure.

The associated risk under decision d and conditional on information state $\mathcal{D}$ is defined as:

$$R(d \mid \mathcal{D}) = \mathbb{E}[C \mid \mathcal{D}, d], \tag{3}$$

The first state of information, $\mathcal{D}_0$ is the information that is available prior to the investigation; it may consist of localized geotechnical information, statistical modelling (assumptions and findings) from previous endeavours [1] and preliminary investigations. It defines the prior probability distributions of all uncertain parameters. In this state of information, the minimum risk for all feasible options is:

$$R_0 = \min_d R(d \mid \mathcal{D}_0), \tag{4}$$

An investigation or monitoring action a generates additional data $\mathcal{D}_a$. Before performing the investigation, the decision maker evaluates the expected minimal risk after observing the yet unknown outcome $\mathcal{D}_a$. This pre-posterior risk is given by [25]:

$$\bar{R}_a = \mathbb{E}_{D_a}\left[\min_d \mathbb{E}[C \mid \mathcal{D}_0, \mathcal{D}_a, d]\right]. \tag{5}$$

The Expected Value of Sample Information (EVSI) is derived by comparing the minimum expected cost with prior information $(\mathcal{D}_0)$ against the expected minimum cost after acquiring new data $(\mathcal{D}_{\text{new}})$:

$$\begin{aligned} \text{EVSI} &= \min_d \mathbb{E}\left[C_{\text{total}}(d \mid \mathcal{D}_0)\right] \\ &-\mathbb{E}_Z\left[\min_d \mathbb{E}\left[C_{\text{total}}(d \mid \mathcal{D}_0, Z)\right]\right], \end{aligned} \tag{6}$$

where the second term represents the pre-posterior analysis, averaged over all possible measurement outcomes Z.

An investigation is economically justified when EVSI exceeds the direct cost of performing the action [12, 26].

The Expected Value of Perfect Information (EVPI) provides a theoretical upper bound on the value of any achievable information and is defined as:

$$\text{EVPI} = R_0 - \mathbb{E}_{\text{true state}}\left[\min_d \mathbb{E}[C \mid \text{true state}, d]\right]. \tag{7}$$

A complete list of the symbols and nomenclature used throughout this paper is provided in Appendix B. These concepts form the basis of an information-based decision-making framework in which risk varies with information. Section 4 demonstrate how spatial random fields and Bayesian Networks can be combined to realize these concepts in the geotechnical area. For a complete mathematical derivation of the expected risk and Value of Information metrics, the reader is referred to Appendix A.

4 Information-value probabilistic risk assessment framework

The information-value probabilistic risk assessment framework is a three-layer, coupled frame- work: a physical framework that describes spatial variability and system behavior/performance, a probabilistic inference layer that defines dependence among the most relevant variables and a decision layer that evaluates risk and information value. Fig. 1 explains that the two-layer coupling is treated as a two-layered combination that was historically assumed to be separate. This allows risk to be expressed as a direct function of information value in the present formulation.

4.1 Physical framework: spatial random fields and performance models

In the physical framework, geotechnical random variable parameters (e.g., cohesion, friction angle, and stiffness) are defined as spatial random fields. These fields are characterized by specified marginal distributions, variances, and spatial correlation structures [4, 5]. Nevertheless, another example is that shear strength parameters are usually assumed to be lognormally distributed. The spatial correlation is generally assumed to be an exponential covariance function of the form:

$$\text{Cov}\{c(s_1), c(s_2)\} = \sigma_c^2 \rho_c(\| s_1 - s_2 \|; \ell_c), \tag{8}$$

where ℓ_c denotes the correlation length. Generation of random field realisations may be accomplished using Fast Fourier Transform spectral methods, turning-bands algorithms or Karhunen–Loève expansions [16].

Wherever there are measurements from the site investigation, unconditional fields are conditioned through kriging or co-kriging. This process certifies that the realized fields honor the measurements but remain spatially covariant [18, 27]. As shown in Fig. 2, such spatially variable parameters have strengths spatially vary within a high and low strength over a distance localized around the critical slip surface and governs its location and shape. Conditioning matters in relation to failure probabilities and inferred failure modes. These spatially variable parameters are then taken to performance based finite element models with strength reductions or calibrated limit equilibrium models. The variables of interest are factors of, for example, factors of safety, displacements or pore pressures. These in turn define the limit state functions that govern each failure mode that can occur. The random fields for cohesion (c) and friction angle (ϕ) are assumed to be lognormally dis-

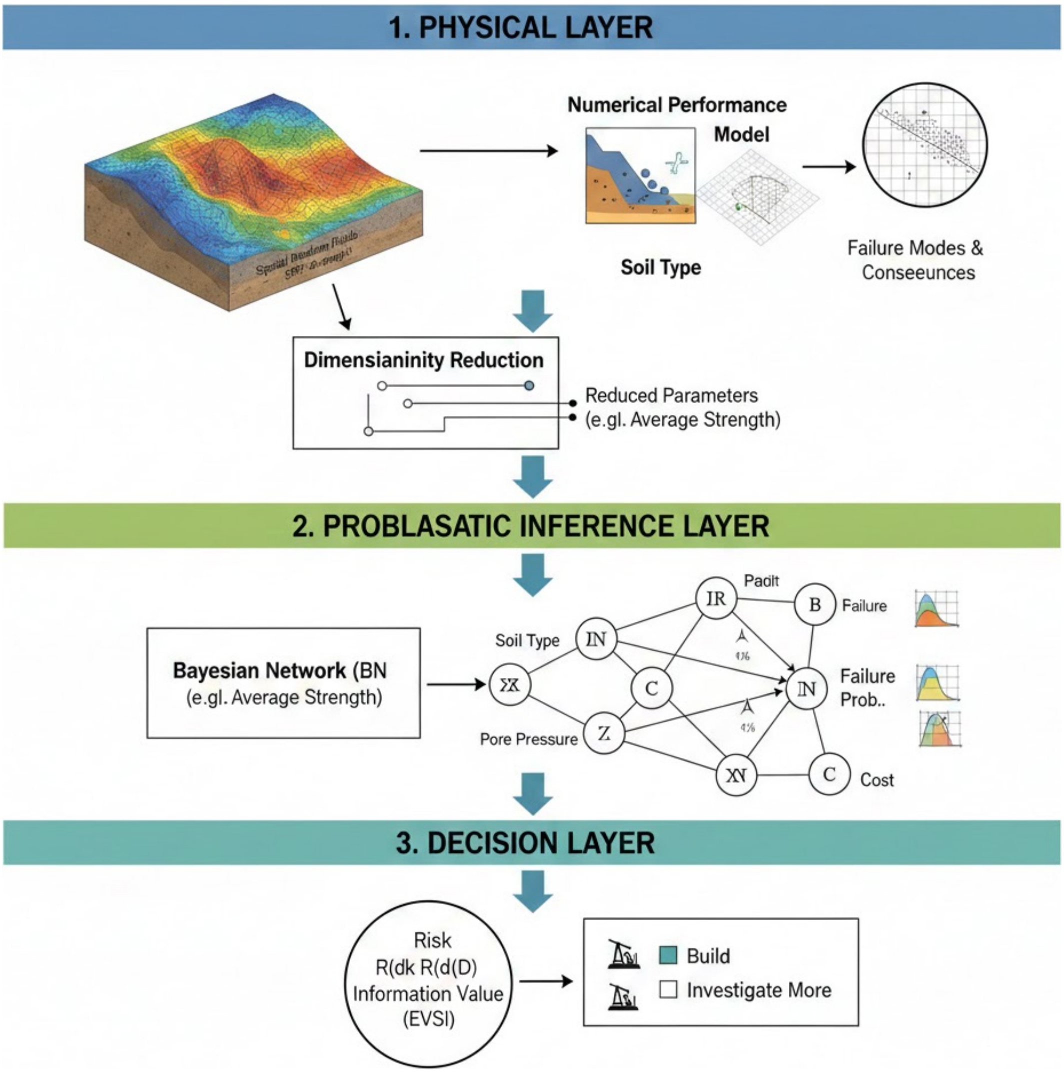

Fig. 1 A schematic representation of the three-layer framework: 1. The Physical-Layer generates spatial random fields; 2. The Inference-Layer assimilates data *via* Bayesian Networks; 3. The Decision-Layer evaluates risk and the Value of Information (VOI)

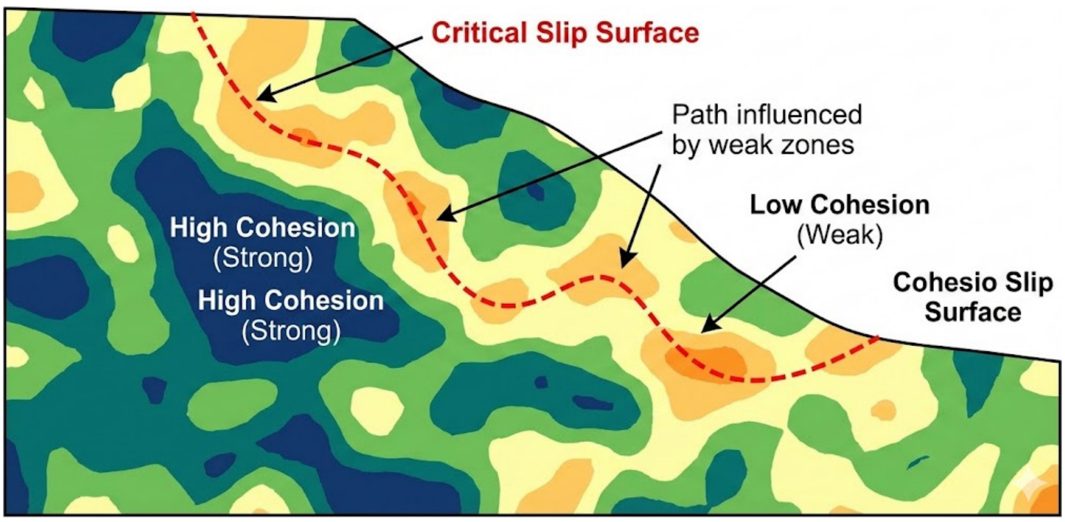

Fig. 2 An illustrative spatial random field of cohesion, highlighting regions of high (*light*) and low (*dark*) strength. The overlay indicates how spatial variability influences the path of the critical slip surface compared to a homogeneous assumption

tributed. The fields are realized from a Gaussian field which is transformed into a standard Gaussian field for conditioning. The transformation is as follows: $Y(x) = \ln(X(x))$. The fields are conditioned by Simple Kriging conditioned on the observations Z. In the transformed space, the measurement error is additive Gaussian noise.

$$Y_{obs} = Y_{\text{true}} + \epsilon,\ \ \epsilon \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right) \tag{9}$$

Equation (9) ensures that the updated variance at the measurement location reflects the measurement uncertainty rather than collapsing to zero.

4.2 Probabilistic inference layer: Bayesian Network representation

Embedding the full high dimensional spatial random field in a Bayesian Network is not feasible with respect to the exponentially large number of conditional probabilities tables. Instead, reduced variables of the random field are obtained. These may be spatially averaged strength parameters applied normal to the slip surface of interest, coefficients of a truncated Karhunen–Loève expansion or performance related variables, say safety factors of trial slip surfaces. A large enough database of simulations is generated by Latin hypercube sampling (LHS) over the random field hyperparameters and the loading parameters. For each realisation the performance model is run, reduced statistics are computed and the outcomes (failure, non-failure) are stored. This dataset constitutes a hard empirical estimate of the conditional probability dependencies that needs to be populated to the Bayesian Network [9, 10, 19, 28].

The nodes in the Bayesian Network take the form of reduced geotechnical input parameters, loading input parameters, observed performance outcomes, failure mechanisms, consequences and observed nodes. The observed nodes are related in a probabilistic fashion to the in-situ measures to allow for measurement errors or measurement noise. The Bayesian Network has the attractive feature of efficiently updating probabilities with the arrival of new evidence and it is transparent in how it disseminates information across it. Fig. 3 illustrates this topological structure of the network and its relation of the reduced variables (cohesion, friction) through to the System Performance (Factor of Safety) and the ultimate Decision Metrics (cost/risk). Additionally, observed nodes permit an update of the belief state upon new in situ data being available.

To formalize the Bayesian Network, we define the observation nodes using a likelihood function those accounts for measurement noise. Let X be the true reduced

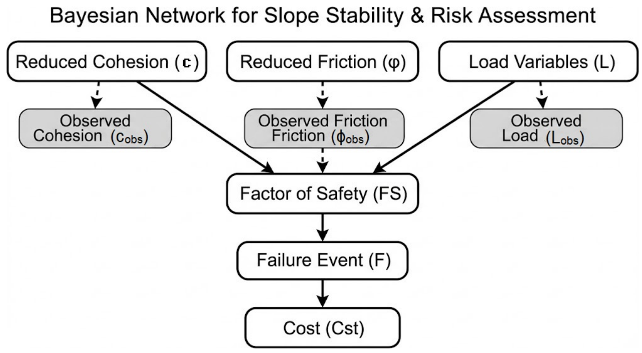

Fig. 3 A representative Bayesian Network (BN) linking reduced geotechnical variables (cohesion (c), friction (ϕ)) and load variables (L) to the Factor of Safety (FS) and Failure Event (F). "Observation" nodes (O_{obs}) allow measurement data to propagate through the network to update failure probabilities

variable (e.g., average cohesion) and Z be the measured value. The probabilistic relationship is modeled as:

$$P(Z \mid X) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left(-\frac{(Z - X)^2}{2\sigma_\epsilon^2}\right), \tag{10}$$

where σ_ϵ represents the standard deviation of the measurement error (noise). For implementation, continuous nodes are discretized into N intervals (states).

For computational implementation, the continuous variables in the Bayesian Network were discretized into $N = 10$ intervals. Through a sensitivity analysis, it was determined that utilizing more than 10 intervals led to negligible changes in the estimated failure probabilities (< 2%). Hence, 10 intervals were chosen for the discretization to maintain reasonable sizes for the conditional probability tables while maintaining accuracy in the prediction.

The conditional probability table (CPT) for a node Y given parents $X_1, \ldots, X_n$ is populated by integrating the joint probability density function over the intervals:

$$P(Y = y_j \mid X = x_i) = \int_{y_j}^{y_{j+1}} f_{Y|X}(y \mid x_i)\, dy. \tag{11}$$

The number of intervals to discretize the continuous variables (N) involves a trade-off between the precision desired and the exponential growth of the size of the CPTs. The size of a CPT with N nodes is of the order $O(N^{n+1})$ where n is the number of parent nodes. The number of intervals used in this work is ($N = 10$). A database of 50,000 data instances was created through the use of Latin Hypercube Sampling (LHS). LHS is a form of sampling that acts as an advanced form of Monte Carlo Simulation (MCS) methods, dividing the distributions and sampling from each to ensure even coverage of the sampling space, thus converging with fewer iterations than Monte Carlo Sampling methods.

During the CPT population, Laplace smoothing with a uniform Dirichlet distribution is used to ensure that no state of any variable is assigned a zero probability. The results of a sensitivity analysis are presented in Table 1. Table 1 shows that (N = 10) results in the lowest mean absolute error of 2.8% when predicting the probability of failure.

Such a discretization process is a trade-off between feasibility and precision; however, it introduces a discretization error term which would have to be regulated by the sensitivity analysis related to the number of bins. A complex geoasset does not fail from one mode of failure alone. A slope fails from deep rotational sliding (a failure assessed deeper) but it can also fail from shallow plane failures or even the failure of the toe bearing capacity. Regarding your comment about combining these probabilities, the Bayesian Network is set up to intrinsically assess "system" reliability. For example, something that should not happen as an independent event, the BN generates a child node called "System Failure" (F_{sys}), the logical union of all other failure mode child nodes (F_1, F_2, …, F_n):

$$F_{sys} = F_1 \cup F_2 \cup \ldots \cup F_n. \tag{12}$$

This hierarchical factorization is justified for two reasons. First, it guarantees that in the case of failure mechanisms, dependencies (such as a spatially continuous weak stratum that promotes sliding and leads to bearing failure) are handled appropriately. Second, it confirms that the model does not give rise to a super-additive excess of consequences. By evaluating the joint probability at the super-element scale, the framework determines the value of one failure mechanism in the existence of another. It proves that the expected cost would be a not super (i.e., marginals independent) conservative approximation of system risk, to which the expected project cost will be assessed.

4.3 Decision layer and Value of Information

The final layer of the framework is the decision layer and transforms these probabilistic estimates into costs of dollars. The objective is to reduce the total anticipated cost over the project's life cycle (incorporating construction, investigation, and potential failure expenses) rather than simply minimizing the failure probability.

4.3.1 Cost function and parametrization

The first step in an objective assessment of design alternatives is a total cost function. Let $d \in \mathcal{D}$ be a given design and investigation effort and let θ be the random realization of the system (i.e., a spatial distribution of soil strengths). The total cost $C_{\text{total}}(d, \theta)$ incurred is the sum of deterministic costs (costs already invested) and stochastic costs (risk costs), as given in Table 2.

Using these definitions, the total cost function is expressed as:

$$C_{\text{total}}(d,\theta) = C_{cons}(d) + C_{inv} + I_F(d,\theta) \cdot C_{\text{fail}}. \tag{13}$$

Taking the expectation over the uncertain state θ, the Expected Total Cost for a strategy is:

$$E\left[C_{\text{total}}(d)\right] = C_{cons}(d) + C_{inv} + P_f(d) \cdot C_{\text{fail}}. \tag{14}$$

4.3.2 Value of Information metrics

The framework employs two key metrics to quantify the economic justification for site investigation:

1. Expected Value of Perfect Information (EVPI): the EVPI represents the theoretical upper bound of what

Table 1 Sensitivity analysis of the BN discretization intervals (N) on model accuracy and CPT size

Number of intervals (N)	Max CPT Size (for 3 parents)	Mean Error in predicted (P_f)	Computational tractability
5	625	14.5%	Very fast
10	10,000	2.8%	Optimal
15	50,625	2.5%	Slow / Sparse data risk
20	160,000	2.4%	Intractable / Overfitting

Table 2 Definition of cost and risk variables

Symbol	Variable Description	Unit	Remarks
$C_{cons}(d)$	Construction cost	Monetary ($)	Deterministic cost associated with a specific design d (e.g., slope angle, reinforcement).
C_{inv}	Investigation cost	Monetary ($)	Cost of acquiring data (e.g., drilling boreholes, laboratory testing). Zero for prior analysis.
C_{fail}	Consequence of failure	Monetary ($)	Aggregated cost of failure, including repair, legal liability, delay penalties, and loss of life (monetized).
$I_F(d, \theta)$	Failure indicator	Binary [0,1]	Function returning 1 if the system fails under state θ and design d; 0 otherwise.
$P_f(d)$	Probability of failure	Probability [–]	Expected value of I_F, calculated *via* the Bayesian Network.

any investigation is worth. It corresponds to the savings achievable if the true state θ were known with certainty before choosing the design. It is defined as the difference between the minimum expected cost under prior uncertainty and the expected minimum cost with perfect knowledge:

$$\begin{aligned}\mathrm{EVPI} = &\min_d E_\theta \left[C_{\text{total}} \left(d, \theta \right) \right] \\ &- E_\theta \left[\min_d C_{\text{total}} \left(d, \theta \right) \right].\end{aligned} \tag{15}$$

2. Expected Value of Sample Information (EVSI): for a realistic site investigation program that yields imperfect data Z (e.g., borehole samples with measurement noise), the EVSI is calculated by performing a pre-posterior analysis. This involves simulating potential measurement outcomes, updating the Bayesian Network for each, and re-optimizing the design:

$$\begin{aligned}\mathrm{EVSI} = &\min_d E_\theta \left[C_{\text{total}} \left(d, \theta \right) \right] \\ &- E_Z \left[\min_d E_{\theta|Z} \left[C_{\text{total}} \left(d, \theta \right) | Z \right] \right].\end{aligned} \tag{16}$$

The expected cost is established by totaling the cost of both decisions (the pre-posterior expected data which was used when making the costs proportional to the uncertain data of Z); therefore, the second term is that of pre-posterior expected cost which is assessed as an expected value based upon the marginal predictive distribution of Z under which EVSI > C_{inv} means that the investigation is cost efficient (i.e., Expected Net Benefit of Sampling (ENBS) > 0). Further details on the pre-posterior analysis and the derivations of both EVPI and EVSI can be found in Appendix A.

5 Application I: example of slope stability

A typical example of slope stability explains the advantage of the framework. This example merges conditional random fields, Bayesian Network updating, and pre-posterior Value of Information (VOI) analysis in a single-problem with a simple geometry. However, the probabilistic dependencies remain complex. The goal is to quantify the transfer of information from the physical layer through the inference layer to the decision layer.

5.1 Geometry and random field specification

The model geometry is a 2D earthen slope of height (H = 10 m and β = 45°). The limit state is governed by effective cohesion (c') and effective friction angle (ϕ'). To model the natural spatial variability, they are defined as stationary, cross-correlated log-normal spatial random fields (SRFs). The parameter values, shown in Table 3, are typical of an archetypical cohesive-frictional soil with high spatial variability [2]. The spatial correlation structure is defined by an exponential autocorrelation function with horizontal (θ_h) and vertical scales of fluctuation of 20 m and 2 m, respectively. This anisotropic structure is representative of naturally occurring layers in sedimentary deposits [4]. A negative cross-correlation between c' and ϕ' is assumed (ρ = −0.5), which is typical for naturally occurring soils [5].

5.2 Performance model and limit state definition

The system performance is assessed using the Factor of Safety (FS) versus rotational failure, evaluated using the simplified Bishop's method. The limit state function (g) is defined as:

$$g = \mathrm{FS} - 1.0. \tag{17}$$

Failure happens when ($g \le 0$). This research study uses a simplified limit-equilibrium procedure for computational proficiency. Nevertheless, the framework is solver-independent and fully suited with advanced Finite Difference Method (FDM) or Finite Element Method (FEM) solvers for composite boundary-value problems.

While this study utilizes the simplified Bishop's method for computational efficiency, the BN architecture is solver-agnostic. The performance node in the network (see Fig. 3) effectively functions as a black-box replacement. Therefore, for complex boundary value problems, this node can be populated using high-fidelity Finite Element Method (FEM) simulations without altering the probabilistic structure of the inference layer.

5.3 LHS database and reduced variable construction

To train the PIM, a LHS Simulation (MCS) database was created. For each realization, the following steps were performed:

Table 3 Summary of geotechnical and probabilistic parameters

Parameter	Symbol	Mean	COV (%)	Distribution	Correlation length (θ_h, θ_v)
Soil unit weight	γ	19 kN/m^3	5%	Lognormal	N/A
Cohesion	c'	20 kPa	30%	Lognormal	20 m, 2 m
Friction angle	ϕ'	30°	10%	Lognormal	20 m, 2 m
Measurement noise	ε	–	10%	Normal	–

1. Random Field Generation: the c' and ϕ' fields are established using the Karhunen–Loève (KL) expansion.
2. Performance Evaluation and Dynamic Mechanism Search: it is a well-known phenomenon in geotechnical engineering that the location of the critical slip surface varies significantly from realization to realization [6, 16]. Thus, no assumptions are made regarding the location of the critical slip surface. For each realization, a dynamic search algorithm determines the minimum factor of safety and the critical slip surface specific to that realization.
3. Dimensionality Reduction: the high-dimensional random fields are reduced to a set of discrete parameters. The spatial averages of the c' and ϕ' values are calculated along the critical slip surface specific to that realization. These spatial averages are referred to as the Reduced Variables (RV_c, RV_ϕ). By extracting these parameters along the unique failure mechanism for each realization, the stochastic nature of the failure mechanisms is represented within the Bayesian Network.

The dimensionality reduction process was successful in collapsing the high-dimensional random fields into a manageable set of discrete nodes for the Bayesian Network [8]. To ensure that the Bayesian Network was truly capable of generalizing the surrogate probabilistic models, a separate hold-out data set was used to validate the Bayesian Network. While the primary database was used to train the Bayesian Network, a separate and independent testing data set comprising 10,000 LHS simulations was used to test and validate the model. The factor of safety estimates from the Bayesian Network were compared to the exact limit equilibrium method calculations from the testing data set. The cross-validation results for this independent data set are shown in Fig. 4. The graph displays an R^2 value of approximately 0.95, indicating that the Bayesian Network was successful in generalizing the surrogate models for the probabilistic relationships between the input and output variables. The mathematical formulation for the functional transformation used to extract these reduced variables is detailed in Appendix A.

5.4 Bayesian Network construction and updating

A Bayesian Network (BN) is constructed to encode the causal relation between the Reduced Variables, the System Performance (FS), and the Failure Event (F). The Conditional Probability Tables (CPTs) are populated using the MCS database [9, 19]. The BN supports bidirectional inference:

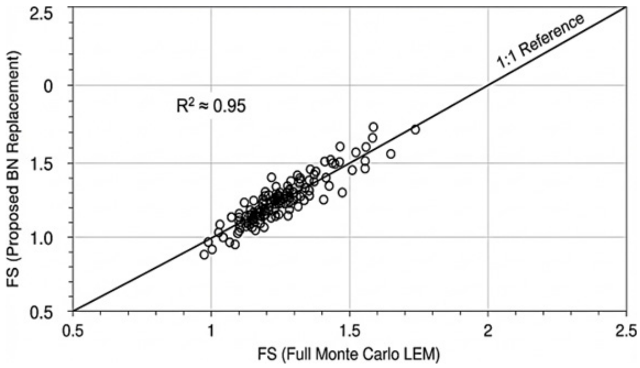

Fig. 4 Comparison of Factor of Safety determined by full Latin Hypercube Sampling (LEM) versus the proposed BN replacement

1. Forward inference, given new soil data, the BN updates (RV_c and RV_ϕ), which propagates to updating the system reliability (P_f).
2. Backward inference, given an observed system performance (slope has not failed), the BN updates the posterior distributions of the soil parameters, hence minimizing the uncertainty.

Initial information about the site is integrated using Kriging, conditioning the random fields on data from two previously drilled boreholes [18], which forms the basis for the upcoming VOI analysis.

5.5 Pre-posterior analysis results

A pre-posterior analysis is conducted to find a hypothetical third borehole that maximizes the Expected Value of Sample Information (EVSI). Virtual boreholes are searched for at intervals of (1 m) along the slope. For each candidate site:

1. The data is sampled, 1000 virtual measurements are drawn from the prior random field model.
2. The Network is updated, each measurement is input to the BN to compute posterior (P_f).
3. The Sampling Cost-Benefit is evaluated, the Expected Net Benefit of Sampling (ENBS) is computed as the expected failure costs minus investigation costs (C_{inv} = 5000).

The analysis reveals a distinct "Value Hotspot" at the toe of the slope. Sampling here yields maximum EVSI, as the slip surface is most responsive to changes in shear strength at the toe, confirming results from previous slope stability investigations [6].

5.6 Comparative analysis: VOI optimization vs. heuristic approach

To assess the efficiency gain provided by the proposed SRF-BN framework, this study compares the optimized investigation strategy against a standard "rule-of-thumb" approach commonly employed in practice. A borehole is often placed at the slope crest ($x = 0$) or mid-slope, typically ignoring spatial correlation. Two courses of action are defined:

- Traditional Course of Action (S_{trad}), drill one borehole at the crest.
- Proposed Course of Action (S_{opt}), drill one borehole at the toe (at EVSI maximum point).

Table 4 confirms that the SRF-BN optimization yields a 143% rise in economic savings (73,000 \$ *vs.* 30,000 \$) compared to the traditional heuristic. This approves that obviously accounting for spatial correlation is critical for cost-effective risk mitigation.

The economics of the proposal is a function of Cost of Failure (C_{fail}). Fig. 5 plots the Expected Net Benefit of Sampling (ENBS) against Cost of Failure. The standard crest borehole (Strad) is economic only when (C_{fail} > \$ 500 k) while the proposed toe borehole (S_{opt}) is valuable for (C_{fail} > \$ 200 k). It is clear that the VOI framework is particularly useful for low-to-medium risk projects which are normally considered uneconomic to sample in the traditional sense.

5.7 Sequential updating and adaptive sampling

A key uniqueness of this procedure is its support for sequential decision-making. By simulating a two-stage adaptive process:

1. Stage 1: the model recognized the toe as the optimal first position. A virtual measurement specified weak soil ($\mu - 1/\sigma$).
2. Update: the BN updated the spatial uncertainty map. The hotspot for information value changed to the mid-slope.
3. Stage 2: the VOI analysis was revaluated. The ENBS for a second borehole at the mid-slope converted positive (+ 15000 \$), supporting further investigation.

On the other hand, if the first borehole specified competent soil, the ENBS for a second borehole became negative. This reveals that the framework can mathematically derive stopping rules for site investigation, simulating the Observational Method [29]. This case study confirms the feasibility of mixing Conditional Random Fields, Bayesian Networks, and VOI analysis into a unified decision-support technique. The results prove that the framework can:

- Identify optimal sampling locations that enlarge risk reduction.
- Quantify the economic advantage of probabilistic planning Surpassing conventional empirical procedures.
- Support adaptive, sequential decision-making.

Table 4 Economic efficiency of investigation strategies

Strategy	Investigation cost (C_{inv}) (\$)	Posterior probability of failure (P_f)	Expected Total Cost (ETC) (\$)	Net benefit (savings) (\$)
Do nothing (Prior)	0	0.120	120,000	–
Traditional (Crest)	5,000	0.085	90,000	30,000
Proposed (VOI - Toe)	5,000	0.042	47,000	73,000

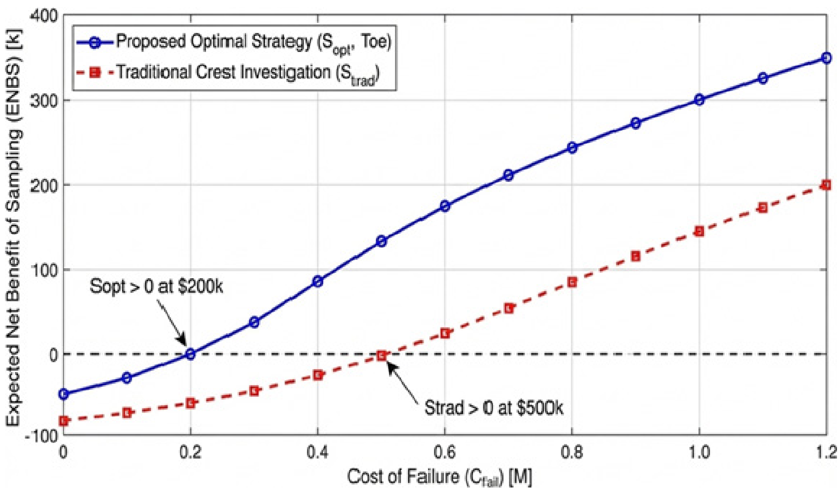

Fig. 5 Sensitivity analysis of the Expected Net Benefit of Sampling (ENBS) relative to the Cost of Failure (C_{fail})

The implementation of a simplified geometry does not constrain the method's applicability to complicated 3D layers; the only limit is the processing power required to train the Bayesian Network.

6 Application II: reliability-based settlement control of shallow foundations

To validate the adaptability of the proposed framework beyond limit equilibrium stability, a second application concentrating on the Serviceability Limit State (SLS) is presented. This case study addresses the risk of extreme differential settlement in a strip footing placing on a spatially variable soil layer.

6.1 Problem definition and spatial variability

The system comprises of a rigid concrete strip footing of width (B = 4 m) exposed to a uniform vertical load (q = 150 kPa). The footing lays on a (10 m) thick clay layer. The governing variable for settlement is the modulus of elasticity (E), which is rendered as a log-normal random field. Contrasting the slope stability case where failure is motivated by a continuous slip surface, settlement is motivated by the volume of soil inside the pressure bulb (stress influence region). The spatial correlation length governs how local soft pockets interact with the rigid-footing:

- Mean (E) = 15 MPa;
- Coefficient of variance (COV) = 40%;
- Correlation length (θ_h = 10 m, θ_v = 1 m).

6.2 Bayesian Network for serviceability

The BN is updated to characterize the stiffness-deformation relationship by:

- Reduced Variables, the random field is minimized to a local average stiffness (E_{avg}) beneath the footing footprint.
- Performance Model, probabilistic settlement (δ) is determined *via* random finite element analysis (RFEM).
- Limit State, failure is defined as total settlement (δ > 50 mm) or differential settlement angular deformation (β > 1/500).

6.3 Comparative VOI analysis

The pre-posterior investigation was repeated to identify the optimal borehole location by:

- Result, contrasting the slope case where the hotspot was at the toe (external to the mass), the EVSI map for the footing reveals the peak value exactly under the foundation center.
- Mechanism, the magnitude of settlement is dominated by the vertically integrated stiffness inside the pressure bulb (1.5 B to 2.0 B depth). Information gathered outside this region has near-zero value because of the vertical scale of fluctuation (θ_v = 1 m) decoupling the stratigraphy.

6.4 Synthesis of applications

The difference between the two case studies highlights the framework's flexibility:

- In slope stability (ultimate limit state), the critical information lies in the shear-band initiation region (Toe).
- In Foundation Settlement (service limit state), the critical information lies in the stress influence region (The Center).

This comparison confirms that the optimal site investigation approach is not static; it is firmly a function of the limit state mechanics and the spatial correlation assembly. The proposed SRF-BN framework appropriately recognizes these distinct optimal solutions without human bias, validating its robustness as a generalized geotechnical decision-support technique.

7 Discussion

The information-based probabilistic framework proposed here is a revolution in geotechnical risk management. Through the integration of SRF, BN and VOI, this research moves beyond estimating the probability of failure to objectively economically incentivizing its reduction. As the applications to both slope stability (ULS) and Shallow Foundation Settlement (SLS) illustrate, the framework addresses the conflict of interest that site investigations pose; the expense of boreholes versus the risk they mitigate over the life of the structure. The contrasting results show that the optimal sampling point shifts from the toe of the slope, where shear strength is dominant, to the center of the foundation, where stiffness is dominant. This confirms that rules of thumb cannot be relied upon in the presence of complex spatial variability. The proposed framework offers ENBS in place of chaotic perception. The risk-benefit ratio of site investigations is quantified. Table 5 summarizes the advantages of this information-driven approach over traditional deterministic and classical probabilistic methodologies.

7.1 Computational efficiency and scalability

A main challenge of pre-posterior analysis is the curse of dimensionality. Discretization of a high-resolution random

Table 5 Comparison of the proposed framework with existing methods

Feature	Traditional deterministic	Standard Probabilistic (PRA)	Proposed information-enhanced framework
Uncertainty Representation	Deterministic (single characteristic values)	Probability distributions (often "lumped" without spatial context)	Spatial Random Fields: explicitly models continuous spatial variability and correlation structures
Risk Quantification	Implicit (Factor of Safety); qualitative risk matrices	Explicit Probability of Failure (P_f); often ignores consequences	Expected Loss: combines (P_f) with failure costs to quantify risk in monetary terms ($)
Information Integration	Static; cannot formally update design with new data	Bayesian updating possible, but often decoupled from spatial geology	Dynamic Updating: uses BNs to propagate specific borehole data through the entire spatial system instantly
Value of Information (VOI)	Qualitative judgment; heuristic planning	Sensitivity analysis (identifies what is important, not where)	Quantitative ENBS: calculates the specific economic value of a borehole before it is drilled
Decision Output	Binary (Pass/Fail)	Reliability Index (β)	Optimized Strategy: identifies the exact location (x, y, z) that minimizes Total Expected Cost

field to a Bayesian Network can lead to an exponential explosion of Conditional Probability Tables (CPTs). This was mitigated in this study by dimensionality reduction using Karhunen–Loève (KL) expansion and path averaging along the most critical failure paths. While successful for the 2D cases presented, this reduction has one drawback: small, local weak zones that do not match the global critical path will be smoothed out. To expand the scope to realistic 3D formations, this challenge could be avoided by using Goal-Oriented Dimensionality Reduction methods, such as Active Subspaces or Sensitivity-Driven KL modes [30]. These methods choose eigenmodes based on which mode maximizes the variance of the safety factor, instead of the random field. This approach ensures the framework focuses on and computes only the features that pose a risk. The computational expense of the nested pre-posterior simulation (inner reliability loop × outer decision loop) was also reduced by using the BN as an instantaneous replacement. Once trained, the BN's inference time is negligible (< 0.1 s), providing real-time decision support in design meetings, which is not feasible with Finite Element LHS methods.

7.2 Addressing deep uncertainty

The case studies presented assumed the statistical hyperparameters of random fields (mean, variance, correlation length) are known. In practice they are uncertain themselves; this is deep uncertainty [1]. If the assumed correlation length is incorrect, the EVSI value calculated will also be incorrect. This can be addressed by extending the framework with Hierarchical Bayesian Models (HBM). Adding a hyperparameter node to the BN would allow it to learn the correlation structure as data is acquired. The VOI of a borehole would then be assessed not only in terms of its ability to measure local strength. It would also be assessed based on its value for minimizing epistemic uncertainty at the global level for the geological model itself. Model-Form Uncertainty (MFU), the discrepancy between the Limit Equilibrium model and reality, could also be accounted for by adding a discrete model node. This permits the network to evaluate the posterior probability of competing physical models (LEM *vs.* FEM), hence ensuring the robustness of decision making.

To achieve the task, the following practical execution steps can perform:

1. Prior Knowledge, establish mean and variance of soil variables from regional archives (D_0).
2. Network Initialization, create the BN structure using the proposed physical-inference coupling (Section 4).
3. Pre-Posterior Check, perform the VOI analysis to calculate if the theoretical (ENBS > 0).
4. Action, if positive, drill at the designated value hotspot (e.g., the Toe).
5. Update, feed the new borehole data into the observation-node to update reliability (P_f).

8 Conclusions

This research study introduced and validated a probabilistic decision strategy for optimizing geotechnical site investigation by combining Spatial Random Fields, Bayesian Networks, and Value of Information analysis. The main drawn conclusions are as follows:

1. The proposed (SRF–BN–VOI) framework presents a practical way to account spatial soil variability in site-investigation decisions. By minimizing high-dimensional random fields into performance-relevant variables, the framework allows Bayesian updating and pre-posterior VOI analysis with no requiring repeated full-scale simulations for every decision alternative.
2. Spatial optimization for sampling is subject to localized variables; it depends on the governing failure

or serviceability mechanism. In the slope stability application, the highest information value was acquired nearby the slope toe, where weak-zones most strongly affected by the critical failure mechanism. In the foundation settlement application, the most valuable sampling position altered toward the center of the foundation, where the stress influence zone governs settlement response.

3. The Bayesian Network surrogate revealed powerful predictive performance in comparison with independent Monte Carlo simulations, with ($R^2 \approx 0.95$). This supports the usage of the minimized-order model for rapid decision support, given that the discretization plan, CPT population, and validation dataset are clearly well-defined.
4. The VOI-based investigation strategy enhanced economic efficiency by up to (143%) compared with traditional heuristic sampling strategies. This proves that site investigation can be treated as an economically justified investment rather than only a regulatory or empirical requirement.
5. The proposed framework is applicable to both ultimate-limit state and serviceability-limit state problems. It can therefore support a wider class of geotechnical decision problems, including slope stability, shallow foundation design, and future monitoring-based digital twin systems.

While this study utilizes assumed random-field hyperparameters and simplified performance models for illustrative purposes, subsequent research should refine this framework. Future efforts ought to implement hierarchical Bayesian modeling to quantify correlation length and hyperparameter uncertainty. Furthermore, accounting for model-form discrepancies across LEM, FEM, and FDM solvers, coupled with validation against established field case studies, will be essential to robustly verify the methodology.

Acknowledgements
The authors would like to acknowledge the great support provided by Mustansiriyah-University, Baghdad-Iraq, which allowed them to achieve this research study.

References

[1] Phoon, K.-K., Kulhawy, F. H. "Characterization of geotechnical variability", Canadian Geotechnical Journal, 36(4), pp. 612–624, 1999.
https://doi.org/10.1139/t99-038

[2] Phoon, K.-K., Kulhawy, F. H. "Evaluation of geotechnical property variability", Canadian Geotechnical Journal, 36(4), pp. 625–639, 1999.
https://doi.org/10.1139/t99-039

[3] Baecher, G. B., Christian, J. T. "Reliability and Statistics in Geotechnical Engineering", John Wiley & Sons, Ltd., 2003. ISBN 0-471-49833-5

[4] Vanmarcke, E. "Random Fields: Analysis and Synthesis", World Scientific Publishing Co Pte Ltd., 2010. ISBN 978-9812563538

[5] Fenton, G. A., Griffiths, D. V. "Risk Assessment in Geotechnical Engineering", Wiley, 2008. ISBN 9780470284704
https://doi.org/10.1002/9780470284704

[6] Cho, S. E. "Probabilistic Assessment of Slope Stability That Considers the Spatial Variability of Soil Properties", Journal of Geotechnical and Geoenvironmental Engineering, 136(7), pp. 975–984, 2010.
https://doi.org/10.1061/(ASCE)GT.1943-5606.0000309

[7] Santoso, A. M., Phoon, K.-K., Quek, S.-T. "Effects of soil spatial variability on rainfall-induced landslides", Computers & Structures, 89(11–12), pp. 893–900, 2011.
https://doi.org/10.1016/j.compstruc.2011.02.016

[8] Papaioannou, I., Straub, D. "Reliability updating in geotechnical engineering including spatial variability of soil", Computers and Geotechnics, 42, pp. 44–51, 2012.
https://doi.org/10.1016/j.compgeo.2011.12.004

[9] He, L., Gomes, A. T., Broggi, M., Beer, M. "Failure Analysis of Soil Slopes with Advanced Bayesian Networks", Periodica Polytechnica Civil Engineering, 63(3), pp. 763–774, 2019.
https://doi.org/10.3311/PPci.14092

[10] Li, X., Zhang, L., Zhang, S. "Efficient Bayesian networks for slope safety evaluation with large quantity monitoring information", Geoscience Frontiers, 9(6), pp. 1679–1687, 2018.
https://doi.org/10.1016/j.gsf.2017.09.009

[11] Raiffa, H., Schlaifer, R. "Applied Statistical Decision Theory", Division of Research, Graduate School of Business Administration, Harvard University, 1961. ISBN 0-87584-017-5 [online] Available at: https://gwern.net/doc/statistics/decision/1961-raiffa-appliedstatisticaldecisiontheory.pdf [Accessed: 05 January 2026]

[12] Wang, X., Shen, C., Xia, M., Wang, D., Zhu, J., Zhu, Z. "Multiscale deep intra-class transfer learning for bearing fault diagnosis", Reliability Engineering & System Safety, 202, 107050, 2020.
https://doi.org/10.1016/j.ress.2020.107050

[13] Straub, D. "Value of information analysis with structural reliability methods", Structural Safety, 49, pp. 75–85, 2014.
https://doi.org/10.1016/j.strusafe.2013.08.006

[14] Gong, W., Tien, Y.-M., Juang, C. H., Martin, J. R., Luo, Z. "Optimization of site investigation program for improved statistical characterization of geotechnical property based on random field theory", Bulletin of Engineering Geology and the Environment, 76(3), pp. 1021–1035, 2017.
https://doi.org/10.1007/s10064-016-0869-3

[15] Rahimian, H., Mehrotra, S. "Distributionally Robust Optimization: A Review", [preprint] arXiv, arXiv:1908.05659, 13 August 2019. https://doi.org/10.48550/arXiv.1908.05659
[16] Griffiths, D. V., Fenton, G. A. "Probabilistic Slope Stability Analysis by Finite Elements", Journal of Geotechnical and Geoenvironmental Engineering, 130(5), pp. 507–518, 2004. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:5(507)
[17] Siacara, A. T., Beck, A. T., Ji, J. "Impact of Random Field Simulations on FEM-Based Earth Slope Reliability", Geotechnical and Geological Engineering, 42(8), pp. 7873–7891, 2024. https://doi.org/10.1007/s10706-024-02956-5
[18] Zhang, G.-H., Chen, W., Jiao, Y.-Y., Wang, H., Wang, C.-T. "A failure probability evaluation method for collapse of drill-and-blast tunnels based on multistate fuzzy Bayesian network", Engineering Geology, 276, 105752, 2020. https://doi.org/10.1016/j.enggeo.2020.105752
[19] Jensen, F. V., Nielsen, T. D. "Bayesian Networks and Decision Graphs", Springer New York, NY, 2007. ISBN 978-0-387-68281-5 https://doi.org/10.1007/978-0-387-68282-2
[20] Straub, D., Papaioannou, I. "Bayesian Updating with Structural Reliability Methods", Journal of Engineering Mechanics, 141(3), 04014134, 2015. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839
[21] Wang, Y., Rezaei, A., Hicks, S. "Digital twin applications in geotechnical engineering: A systematic review", Machine Learning and Data Science in Geotechnics, 1(1), pp. 78–92, 2025. https://doi.org/10.1108/MLAG-02-2025-0005
[22] Cotoarbă, D., Straub, D., Smith, I. F. C. "Probabilistic digital twins for geotechnical design and construction", [preprint] arXiv, arXiv:2412.09432, 12 December 2024. https://doi.org/10.48550/arXiv.2412.09432
[23] Strong, M., Oakley, J. E., Brennan, A. "Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: A nonparametric regression approach", Medical Decision Making, 34(3), pp. 311–326, 2014. https://doi.org/10.1177/0272989X13505910
[24] Heath, A., Manolopoulou, I., Baio, G. "Efficient Monte Carlo Estimation of the Expected Value of Sample Information Using Moment Matching", Medical Decision Making, 38(2), pp. 163–173, 2018. https://doi.org/10.1177/0272989X17738515
[25] de Rocquigny, E., Devictor, N., Tarantola, S. (eds.) "Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management", John Wiley & Sons, Ltd., 2008. ISBN 9780470994474 https://doi.org/10.1002/9780470770733
[26] Van Coile, R., Caspeele, R., Taerwe, L. "Cost-Optimization of Concrete Slabs Exposed to Natural Fires, Taking into Account Post-Fire Behavior", In: Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM) and the Sixth International Symposium on Uncertainty, Modeling, and Analysis (ISUMA), Liverpool, UK, 2014, pp. 194–203. ISBN 9780784413609 https://doi.org/10.1061/9780784413609.020
[27] Wang, Z., Yu, Q., Wang, J., Hu, Z., Wang, A. "Grammar Correction for Multiple Errors in Chinese Based on Prompt Templates", Applied Sciences, 13(15), 8858, 2023. https://doi.org/10.3390/app13158858
[28] Lakermi, A., Allal, A. M. "Application of Bayesian Networks for Geotechnical Risk Analysis: Case Study in Algeria", Geotechnical and Geological Engineering, 39(2), pp. 815–828, 2021. https://doi.org/10.1007/s10706-020-01523-y
[29] Peck, R. B. "Advantages and Limitations of the Observational Method in Applied Soil Mechanics", Géotechnique, 19(2), pp. 171–187, 1969. https://doi.org/10.1680/geot.1969.19.2.171
[30] Constantine, P. G., Dow, E., Wang, Q. "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces", SIAM Journal on Scientific Computing, 36(4), pp. A1500–A1524, 2014. https://doi.org/10.1137/130916138

Appendix A

Appendix A summarises key mathematical derivations underlying the information-enhanced probabilistic framework.

Let $\mathcal{D}$ denote the available information set. The risk associated with decision d is defined as the expected total cost conditioned on this information:

$$R(d \mid \mathcal{D}) = E_{\theta \mid \mathcal{D}}\left[C_{\text{total}}(d,\theta)\right],$$

where the total cost C_{total} is decomposed into construction, investigation, and monitoring costs, together with failure consequences:

$$C_{\text{total}}(d,\theta) = C_{cons}(d) + C_{inv}(\mathcal{D}) + C_{\text{fail}} \cdot I_F(d,\theta).$$

Here, $I_F(d, \theta)$ is the binary failure indicator (1 if failure occurs, 0 otherwise). Expanding the expectation yields:

$$R(d \mid \mathcal{D}) = C_{cons}(d) + C_{inv}(\mathcal{D}) + C_{\text{fail}} \cdot P(F \mid d,\mathcal{D}).$$

Before execution an investigation action (a), the measurement outcome (Z) is unknown. The pre-posterior anticipated minimal risk is determined by averaging the posterior risks over all possible measurement products:

$$R_a = E_Z\left[\min_d R(d,\mathcal{D}_a)\right].$$

Using the law of total expectation, the difference between the current minimal risk (R_0) and the expected future minimal risk defines the Expected Value of Sample Information (EVSI):

$$\text{EVSI} = R_0 - R_a,$$

where:

$$R_0 = \min_d R(d,\mathcal{D}_0).$$

If the true state of the system θ were perfectly known, the optimal decision would satisfy:

$$C_{opt}(\theta) = \min_{d} C_{\text{total}}(d,\theta).$$

The corresponding expected risk with perfect information (R_{PI}) is:

$$R_{\text{PI}} = \mathbb{E}_{\theta}\left[C_{\text{total}}(d,\theta)\right].$$

Thus, the Expected Value of Perfect Information (EVPI) is:

$$\text{EVPI} = R_0 - R_{\text{PI}}.$$

Since perfect information is unattainable in practice, EVPI provides an upper bound on the economic value of any feasible investigation programme.

Let $X(s)$ be a random field defined over domain Ω. A reduced variable Z_r may be obtained by a functional transformation:

$$Z_r = \int_{\Omega} X(s)\, w(s)\, ds,$$

where $w(s)$ is a weighting function representing the influence of each spatial location on the failure mechanism (e.g., shear strain localisation).

Alternatively, using a truncated Karhunen–Loève (KL) expansion:

$$X(s) \approx u_X(s) + \sum_{k=1}^{m} \sqrt{\lambda_k}\, \phi_k(s)\, Y_k,$$

where the reduced variables are the coefficients $Y_1, \ldots, Y_m$, with m chosen to retain the dominant variance modes of the field.

Appendix B

Table B1 The principal symbols used throughout the paper

Symbol	Meaning
$X(s)$	Spatial random field of geotechnical properties at location s
θ	Hyperparameters governing random field structure (means, variances, correlation lengths)
L	Uncertain load variables (e.g., groundwater level, surcharge)
g_j	Limit state function for failure mode j
F_j	Failure event $g_j \leq 0$
F_{sys}	System failure event (union of individual failure modes)
C_j	Consequence (cost) associated with failure mode j
C_{total}	Total cost including construction, investigation, monitoring and failure losses
d	Design or mitigation decision
$\mathcal{D}_0$	Initial (prior) information set
$\mathcal{D}_a$	New information set obtained from investigation action a
$R(d \mid \mathcal{D})$	Expected risk under decision d and information $\mathcal{D}$
R_0	Minimal expected risk under prior information
R_a	Pre-posterior expected minimal risk after action a
EVSI	Expected Value of Sample Information
EVPI	Expected Value of Perfect Information
Z	Reduced variable extracted from a spatial random field (or measurement outcome)
μ_X	Mean of random field X
$\phi_k(s)$	k-th eigenfunction in Karhunen–Loève (KL) expansion
λ_k	k-th eigenvalue in KL expansion
Y_k	Random coefficients in reduced random field approximation