OPTIMIZATION OF GROUND INVESTIGATIONS WITH THE VALUE OF INFORMATION ANALYSIS

Ivan Depina[1] and Rui Tao[1]

KEYWORDS

Value of Information, Ground, Investigations, Optimization

ABSTRACT

This study investigates the problem of optimizing the number and locations of the ground investigations. Determining the number and locations of geotechnical investigations is central in reducing the ever-present uncertainties in soil properties. Good knowledge and understanding of the soil properties is central in ensuring that the design meets necessary safety and cost constraints. The selection of the numbers of ground investigations and their locations often varies from site to site due to lack of objective measures for determining the optimal ground investigation program. This results in a great variation of the numbers and locations of ground investigations in practice. This study aims to approach the problem from a probabilistic perspective and propose a more explicit and objective framework for the determination of the optimal numbers and locations of ground investigations. The proposed framework is based on a probabilistic geotechnical modelling approach and the Value of Information (VoI) analysis. The value of ground investigations results from the improved knowledge and the reduction of the potential negative consequences in case of an inadequate design. The potential benefits of conducting additional soil investigations are compared to the costs of ground investigations to determine the optimal number of samples and locations. The performance of the proposed algorithm is investigated on a slope stability problem.

1. INTRODUCTION

Geotechnical engineers are often faced with the task of making decisions under uncertainties resulting from lack of knowledge about ground properties, loads, soil behavior etc. In making these decisions, they are commonly required to balance the competing goals of safety and costs without having an explicit approach in dealing with the decision-making problem. This often results in a

[1] Norwegian University of Science and Technology, Department of Civil and Environmental Engineering

great variation in the conduced number of ground investigations and their locations across sites and projects. This study aims to examine the decision-making problem in a more explicit way by casting the problem in a probabilistic setting and applying the Value of Information (VoI) analysis to determine the optimal number and locations of ground investigations.

2. CONDITIONAL RANDOM FIELD

Two-dimensional conditional Gaussian random field is implemented to model uncertain soil properties and update them with measurements [e.g., 1]. Measurements of soil properties include measurement and interpretation errors. A discretized random field is denoted as $R(x, z)$ and specified with a vector of mean values $\mu = [\mu(x_1, z_1), \ldots, \mu(x_m, z_m)]^T$ and a covariance matrix Σ with the following elements $\Sigma_{i,j} = \sigma(x_i, z_i)\sigma(x_j, z_j)\rho\left((x_i, z_i), (x_j, z_j)\right);\ i, j = 1, \ldots, m$, where σ is the standard deviation and ρ is the correlation coefficient modelled with an ellipsoidal autocorrelation model parameterized with horizontal and vertical correlation lengths, θ_x and θ_z, respectively [e.g., 1]. The joint probability density function (pdf) of R corresponds to a multivariate normal pdf, defined as follows:

$$R \sim \mathrm{f}(\mathbf{r}) = \mathrm{N}(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Consider that observations or measurements are available on s locations of the discretized domain, $\boldsymbol{o} = \{o_1, \ldots, o_n\}$ with the covariance matrix Σ_o. The likelihood of the observations is:

$$\boldsymbol{o} \sim f(\boldsymbol{o}) = N(\boldsymbol{o}; \boldsymbol{H}\mu, \boldsymbol{\Sigma_o})$$

where H is a linear function, specified as a matrix composed of s rows and m columns. Each row of the H matrix is assigned to an observation with the value of one on the column index corresponding to the index of the observation. The rest of the elements have zero values. The joint pdf of the random field conditioned on the observation is defined as [e.g., 1]:

$$[R|\boldsymbol{o}] \sim f(r|\boldsymbol{o}) = N\left([r|\boldsymbol{o}]; \mu_{r|\boldsymbol{o}}, \Sigma_{r|\boldsymbol{o}}\right)$$

where

$$\mu_{r|\boldsymbol{o}} = \mu + \boldsymbol{\Sigma H}^T[\boldsymbol{H\Sigma H}^T + \boldsymbol{\Sigma}_o]^{-1}(\boldsymbol{o} - \mu\boldsymbol{H})$$

$$\boldsymbol{\Sigma}_{r|\boldsymbol{o}} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathrm{H}^\mathrm{T}[\boldsymbol{H\Sigma H}^T + \boldsymbol{\Sigma}_o]^{-1}\,\boldsymbol{H\Sigma}$$

with $\mu_{r|\boldsymbol{o}}$ being the mean, and $\boldsymbol{\Sigma}_{r|\boldsymbol{o}}$ the covariance matrix of the conditional random field.

3. VALUE OF INFORMATION ANALYSIS

The decision model presented in [2] is adopted in this report. A binary variable Y is introduced to indicate the performance of a system with $Y = 1$ denoting failure and $Y = 0$ denoting non-failure [e.g., 2]. In case that the system is not able to satisfy the design requirements, a set of measures can be applied to improve the system response and satisfy the design requirements. Let d denote a decision variable with $d = 0$ denoting no measures and $d = 1, \dots, n_d$ the d^{th} decision for a measure improving the system performance, where n_d is the number of alternative engineering measures [e.g., 2]. Losses due to failure of a system can include economic, environmental, and health consequences and are denoted with c_f, while the costs of engineering measures are denoted as $c_e(d)$. Furthermore, the following restrictions are made on the costs, $c_e(0) = 0$ and $c_e(d) > 0$ of $d > 0$ [e.g., 2].

The decision-making process is based on a utility function, $u(Y, d)$, that encapsulates the major considerations including safety and costs of engineering measures. The following utility function is selected in this study [e.g., 2]:

$$u(Y, d) = \begin{cases} -c_e(d), & \text{if } Y = 0 \\ -\left(c_e(d) + c_f\right), & \text{if } Y = 1 \end{cases}$$

The likelihood of a geotechnical system experiencing safe performance or failure depends on the uncertain and spatially variable soil properties, R, which are modelled with the conditional random field approach introduced earlier. Let $\mathrm{f}(\mathbf{r})$ denote the prior knowledge about uncertain soil properties available before conducing soil investigations. Realizations from the prior distribution are used as an input to geotechnical model to evaluate the system response in terms of a limit state described by the performance function $g(\mathbf{r}, d)$, for a given decision variable d. The prior failure probability is evaluated as follows [e.g., 2]:

$$p_{f,0} = P(Y = 1 | d = 0) = \int_{g(\boldsymbol{r},d) \leq 0} f(\boldsymbol{r}) d\boldsymbol{r}$$

The expected benefit of taking a decision d can be calculated as the expectation of the utility function [e.g., 2]:

$$E_Y[u(Y, d)] = u(0, d)\left(1 - p_{f,0}\right) + u(1, d) p_{f,0} = -\left[c_e(d) + c_f p_{f,0}\right]$$

where $E_Y[\cdot]$ denotes expectation with respect to the distribution of Y [e.g., 2]. The expected benefit will be used as a basis for deciding among the alternatives, with the optimal decision being the one that maximized the expected outcome.

Initially, the expected utility is calculated based on prior information on the uncertain parameters $\boldsymbol{r}$. After performing field or laboratory investigations, the knowledge on the uncertain parameters can be updated with observations, $\boldsymbol{O}$, based on the Bayes' theorem. In this study both the prior and the likelihood are multivariate normal distributions. This allows for analytical solution for the Bayesian updating problem. The posterior failure probability is denoted as $p_{f,d} = P(Y = 1 | \boldsymbol{O}, d)$. The posterior failure probability can be used to update the expected utility as follows [e.g., 2]:

$$E_{Y|\boldsymbol{O}}[u(Y,d)] = u(0,d)\left(1 - p_{f,d}(\boldsymbol{O})\right) + u(1,d)p_{f,d}(\boldsymbol{O})$$
$$= -\left[c_e(d) + c_f p_{f,d}(\boldsymbol{O})\right]$$

where $E_{Y|D}[\cdot]$ is the expectation with respect to the updated distribution of Y given $\boldsymbol{O}$. One may be able to update their decision based on the updated utility function. The difference between the maximum expected utility under the optimal decisions before and after collecting observations $\boldsymbol{O}$ is called Conditional Value of Information (CVOI) [e.g., 2]:

$$CVOI = max_d E_{Y|\boldsymbol{O}}[u(Y,d)] - max_d E_Y[u(Y,d)]$$

CVOI measures the improvement of the expected utility after collecting additional information from field or laboratory investigations. Prior to performing field or laboratory investigations, it is important to assess the effectiveness of the planned investigation program. Before the investigation program is conducted it is necessary to evaluate the uncertainty in observations $\boldsymbol{O}$ based on the prior knowledge. Let $f_{IP}(\boldsymbol{O})$ denote the pdf of $\boldsymbol{O}$ from a given Investigation Program (IP) that is inferred from the prior knowledge. Mathematically, it can be inferred from the total probability theorem as follows [e.g., 2]:

$$f_{IP}(\boldsymbol{O}) = \int f_{IP}(\boldsymbol{O}|\boldsymbol{r})f(\boldsymbol{r})d\boldsymbol{r}$$

The potential benefit of performing the IP is evaluated by conducting Value of Information (VoI) analysis. VoI is defined as the average of CVOI considering all possible values of $\boldsymbol{O}$ [e.g., 2]:

$$VoI_{IP} = \int CVOI_{IP}(\boldsymbol{O})f_{IP}(\boldsymbol{O})d\boldsymbol{O}$$

Integrating the earlier expression into the value of information analysis for the considered IP, VoI_{IP}, the following expression can be obtained [e.g., 2]:

$$VoI_{IP} = \int \max_d \{-\left[c_e(d) + c_f p_{f,d}(\boldsymbol{O})\right]\} f_{IP}(\boldsymbol{O})\, d\boldsymbol{O} - \max_d \{-[c_e(d) + c_f p_{f,0}]\}$$

4. RESULTS AND DISCUSSION

Geometry of the slope stability problem is shown in Figure 1. The slope is 20 m high with the slope inclination of 1:2. Slope stability is analyzed with the Janbu's direct method, $F_S = \bar{s}_u \frac{N_0}{P_d} + \varepsilon_M$, where $\bar{s}_u$ is the average shear strength along the failure surface, N_0 is the stability coefficient, and $P_d = \gamma \cdot H$, where $\gamma = 19kN/m^3$ and $H = 20\ m$, and $\varepsilon_M \sim N(0{,}0.05)$ is the assumed model error. Reliability analysis is conducted with the Monte Carlo method to evaluate the probability that the factor of safety is lower than one, $P_F = P(F_S \leq 1) =$

$P\left(\bar{s}_u \frac{N_0}{P_d} + \varepsilon_M \leq 1\right)$. F_S is calculated by considering 12 potential failure surfaces ending in the proximity of the slope toe. The surface with the lowest factor of safety is selected as the critical failure surface.

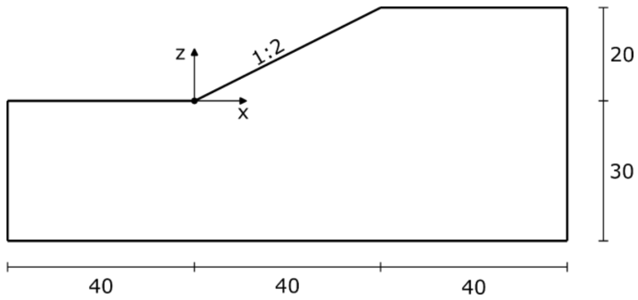

Figure 1 Slope stability problem geometry.

Prior knowledge on undrained shear strength is specified with a normal random field defined with a linearly increasing trend, $\mu_{s_u} = 40 + 1.8z$, constant standard deviation, $\sigma_{s_u} = 8\ kPa$, and correlation lengths of $\theta_x = \theta_z = 50\ m$. The actual and, initially unknown, soil properties are generated from a realization of a normal random field defined with a linearly increasing trend, $\mu_{s_u} = 40 + 1.8z$, constant standard deviation, $\sigma_{s_u} = 1.2\ kPa$, and correlation lengths of $\theta_x = \theta_z = 50\ m$. VoI analysis is conducted based on a utility function introduced earlier with the failure cost being $c_f = 1$ and the cost of soil investigation being $c_e = 0.01$. VoI analysis was conducted to simulate the procedure for performing 10 soil investigations.

Initially, prior analysis was conducted to assess the safety of the slope based on the prior knowledge. Initial failure probability was estimated to $P_{F0} = 0.119$. VoI analysis was the conducted to simulate 10 soil investigations. Prior to performing soil investigation, a trial analysis was conducted. In the trial analysis, performing soil investigations was simulated at all the points in Figure 2, while accounting for the inherent variability and measurement errors. VoI values after the trial analysis for the first investigation are shown in Figure 3. Positive values show that that there is a benefit in performing investigation at that location. The optimal location is identified as the location that maximizes the expected VoI. Soil investigation is then performed at the optimal location, and the measured value with the addition of the measurement error, $\epsilon \sim N(0{,}0.05)$, is used to update the mean and standard deviation of the conditional random field.

Figure 4 shows the optimal locations for the first five investigations and the effects of investigations on the uncertainties in soil properties. Performing ground investigations reduces the original standard deviation of 8 kPa to the inherent variability and measurement error. The reduction of uncertainty does not only affect the location of the measurement, but also the neighboring areas due to spatial similarity, simulated with the autocorrelation function. Figure 3 shows also that the VoI algorithm was able to independently identify points along the potential failure surfaces as the optimal points for performing investigations.

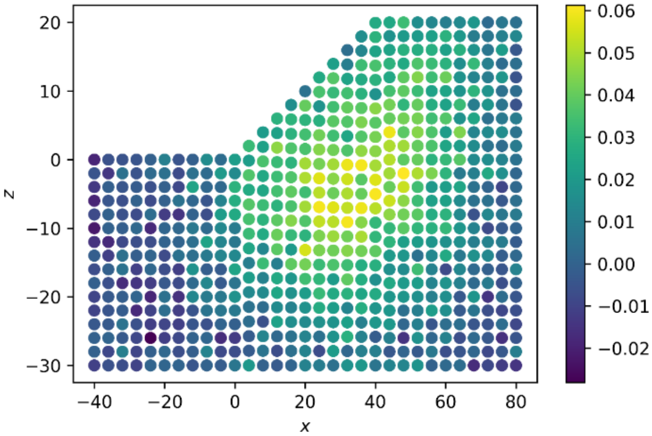

Figure 2 VoI values after trial analysis before for the first investigation.

Figure 4 compares the VoI of performing the 10 investigations and the costs of investigations. Although VoI is higher than the cost of investigations for all the 10 soil investigations, one can see that after the first 4 investigations, the increase in the VoI is not significant anymore. The gains in VoI from performing additional soil investigations are diminishing, which can be also seen in Figure 5 that shows the VoI subtracted for the cost of investigations. Thus results in 4 ground investigations as the optimal number of ground investigations.

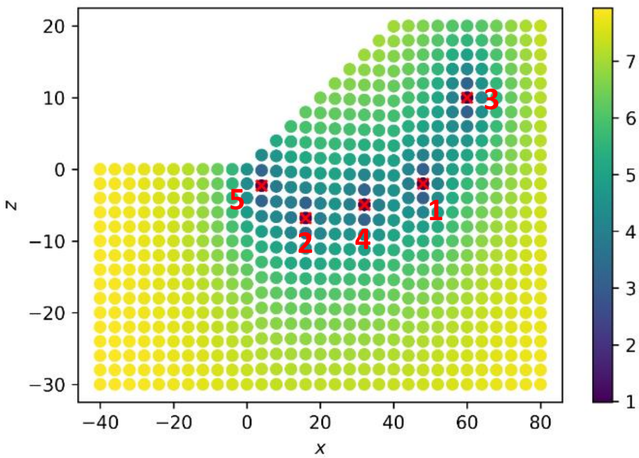

Figure 3 Effects of measurement on soil uncertainties and the optimal locations of the first five ground investigations.

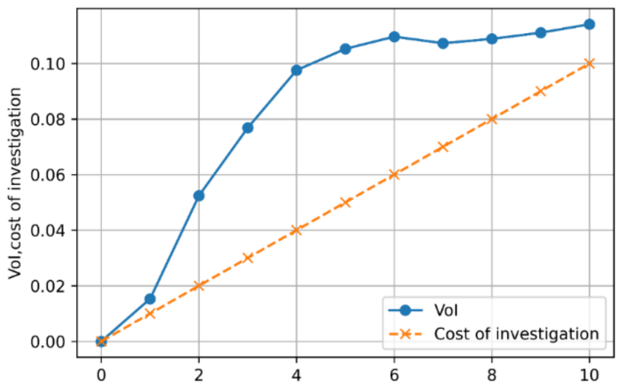

Figure 4 Comparison of VoI and costs of ground investigations.

Additional results were conducted for a range of correlation lenghts, with the results presented in Table 1. Higher values of correlation lengths have a positive effect on VoI as a measurement leads to reduction in uncertainties in a larger area surrounding the measurement location. This resulted in a larger number of soil investigations being more optimal.

Conversely, low values of correlation lengths resulted in low VoI values because low correlation length resulted in the impact of the measurement being limited to a smal zone around the soil measurement. Additionally, due to relatively coarse discretization of the domain in the slope stability problem, the estimates of factor of safety did not benefit substantially from those measurements. Implementation of more advanced slope stability models with a finer discretization and non-circular failure surfaces is needed to better asses the effects of small correlation lengths on the optimal number of soil investigations. Determining realistic values of correlation lenghts is important for assessing the optimal number of samples based on the implemented VoI methodology.

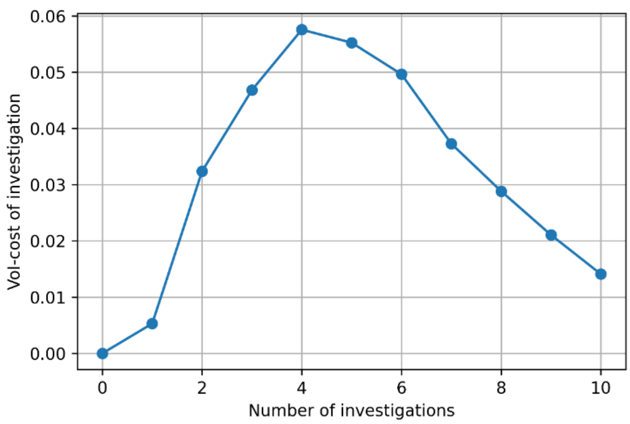

Figure 5 VoI subtracted for the costs of ground investigations.

Table 1: Optimal numbers of ground investigations for varying correlation lengths.

θ_x	θ_z	N_{Opt}
100	2	2
100	10	4
10	10	0
50	50	4
100	100	3
1000	10	3
50	20	4
100	2	2

5. CONCLUSIONS

The report demonstrated the use of Value of Information (VoI) analysis in planning soil investigations and determining the optimal number of soil investigations. VoI analysis was combined with a conditional random field model that allowed integration of soil investigations, with the addition of measurement and interpretation errors, in the modelling of uncertain soil properties.

The results show that VoI analysis can manage the trade-off between cost of investigation and potentially adverse consequences of geotechnical system failure to determine the optimal number of soil investigations and their locations. The results showed that the optimal number of investigations is significantly affected by the correlation length. Higher correlation length allows a measurement to have a greater impact on the reduction of uncertainties in the surrounding soil zone, which results in a greater VoI and bigger increase in safety in comparison to situations with lower correlation lengths.

ACKNOWLEDGEMENT

The author acknowledges the financial support from the Norwegian Public Road Authority.

REFERENCES

[1] Ivan Depina and Amanuel Petros Wolebo. Application of conditional random fields for the reliability assessment of slope stability. In 12th International Conference on Structural Safety & Reliability, 2017

[2] JZ Hu et al. Value of information analysis of site investigation program for slope design. Computers and Geotechnics, 131:103938, 2021