Geotech Geol Eng
https://doi.org/10.1007/s10706-023-02403-x

ORIGINAL PAPER

Optimal Site Investigation Through Combined Geological and Property Uncertainties Analysis

Opeyemi E. Oluwatuyi · Kam W. Ng · Shaun S. Wulff · Rasika Rajapakshage

Received: 12 December 2022 / Accepted: 17 February 2023

Abstract Site investigation is crucial in characterizing the geomaterial profile for the design of bridge pile foundations. A site investigation plan should be conducted to maximize geomaterial information and minimize uncertainty. Thus, both geological and property uncertainties should be explicitly incorporated into a site investigation plan. This leads to the question of how to choose the corresponding optimal number and location of boreholes in a multiphase site investigation plan in order to reduce these uncertainties. This study addresses these problems using multinomial categorical prediction and universal kriging on a random field with multiple simulations. Site investigation data for this study are taken from a bridge project in Iowa, USA, which consists of four boreholes, each within the proximity of the pile foundation location. Subsequent numbers of recommended boreholes and their associated locations are determined to minimize the combined uncertainties. The effectiveness of this combined analysis for determining an optimal site investigation plan (OSIP) is validated and compared to an analysis done solely on property uncertainty. The proposed OSIP yields a lower prediction error, improves the prediction of geomaterial type and property, and reduces the subsurface uncertainties. The incorporation of OSIP invariably improves the design efficiency and performance of bridge pile foundations.

Keywords Geostatistics · Geological uncertainty · Property uncertainty · Site investigation · Simulation

Abbreviations

AIC	Akaike information criteria
AICc	Akaike information criteria corrected
BH	Borehole
BIC	Bayesian information criteria
IADOT	Iowa department of transportation
IGM	Intermediate geomaterial
LOOCV	Leave one out cross validation
MCP	Multinomial categorical prediction
MCS	Monte Carlo simulation
n2llik	Loglikelihood of normal distribution
OSIP	Optimal site investigation plan

O. E. Oluwatuyi · K. W. Ng (✉)
Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, 1000 E. University Ave, EN3050, Laramie, WY 82071-2000, USA
e-mail: kng1@uwyo.edu

O. E. Oluwatuyi
e-mail: ooluwatu@uwyo.edu

S. S. Wulff
Department of Mathematics and Statistics, University of Wyoming, 1000 E. University Ave, Laramie, WY 82071-3036, USA
e-mail: wulff@uwyo.edu

R. Rajapakshage
Department of Computer Systems Engineering, Faculty of Computing and Technology, University of Kelaniya, Kelaniya, Sri Lanka
e-mail: rasikar@kln.ac.lk

Published online: 27 February 2023

PU	Property uncertainty
PUGU	Property uncertainty and geological uncertainty
RMSE	Root mean square error
SD	Standard deviation
spMC	Spatial Markov chains
SPT	Standard penetration test
UK	Universal kriging

1 Introduction

Site investigation is an essential aspect of geotechnical design. Two main classes of geomaterial-related information are obtainable from the sampling aspect of site investigation, namely the (1) geomaterial property where the data consists of a quantitative geotechnical response, and (2) geomaterial type where the data consists of categorical geomaterial classification. Hence, an adequate geospatial model requires the specification of these two uncertainties (Bock 2006). Property uncertainty refers to uncertainty due to the prediction of a geomaterial property at the unsampled locations of the site. It is closely related to inherent variability which is the difference between a geomaterial property from one spatial location to another (Liao et al. 2022; Mendoza and Hurtado 2022; Zhang et al. 2023). Geological uncertainty is the changeability due to one geomaterial embedded in another, especially at the boundaries between different geomaterial layers (Oluwatuyi et al. 2023a, b). To obtain an optimal site investigation plan (OSIP), emphasis must be placed on reducing both forms of uncertainties. Moreover, an OSIP obtained from the analysis of both forms of uncertainties will result in a better prediction of the geomaterial property for pile foundation design (Juang et al. 2019). However, geological uncertainty has been ignored in previous OSIP studies (Abdulai and Sharifzadeh 2019; Oluwatuyi et al. 2022a; Zhang et al. 2022). In the risk assessment analysis for soft-ground tunneling, the influence of geological uncertainty is integrated into the spatial modeling of geomaterial property uncertainty to aid in engineering decision-making (Grasmick et al. 2020). Accurate characterization of the geomaterial boundaries (geological uncertainty) using probabilistic tools can result in a better evaluation of slope reliability (Deng et al. 2017) and landslide stability (Wang et al. 2018; Gong et al. 2020).

Another difficulty in the geotechnical characterization of a site is the determination of the optimal number and location of boreholes specifically for a bridge project. To conduct a site investigation, one needs to consider the: (1) borehole depth; (2) the number of drilled boreholes; (3) the locations and spacings of the boreholes to be drilled, for a thorough and well-organized site investigation (Jelušič and Žlender 2014). In practice, site investigation plans are based on budget, local experience, or procedures specified by codes. For example in the United States, AASHTO (2020) specifies a minimum of two boreholes per bridge pier or abutment for a bridge with a width greater than 30 m. In Australia, the recommendation of a borehole at every 30 m is specified by the Austroads Bridge Design Code (Wedgwood 1992). The summary of these open-ended design codes is a specification of uniformly spaced boreholes that are not directly based upon uncertainty or cost-efficiency. A site investigation plan based on the design code, or one determined empirically on time, budget, and experience can be inefficient because it does not account for prior information from the specific project site. The optimal number of boreholes on site A might be inappropriate or drilled at the wrong locations for site B. Hence, site-specific determination of the optimal number and location of boreholes for the design of bridge pile foundations remains a largely unsolved problem. The use of geostatistical/computational approaches to predict geomaterial information, quantify its uncertainties, and optimize site investigation had been discussed in recent studies. Efficient sampling locations for one-dimensional site characterization are determined by using information entropy and Bayesian compressive sampling (Zhao and Wang 2019). The accuracy of identifying and drilling horizontal wells from high-quality marine shale gas reservoirs is improved upon by Zhu et al. (2021) through the use of the oversampling method and random forest algorithm. A general guide to optimizing site investigations for pile design in a single layer of soil was proposed by Crisp et al. (2020). The study by Shi and Wang (2021) focused on the site's geological uncertainty as obtained through multiple-point statistics and information entropy in the determination of borehole numbers and locations for slope stability analysis. Žlender et al. (2012) predicted, optimized, and spaced borehole drillings by using artificial intelligence techniques. However,

optimization in these recent studies is either restricted to the property or geological uncertainty of the site. Both geological and property uncertainties should be addressed concurrently (Boumezerane et al. 2014; Han et al. 2020; Mazraehli and Zare 2022). Thus, this present study presents a combined analysis of both uncertainties to optimize the site investigation plan.

This study is aimed at optimizing the site investigation plan by interpolating the subsurface stratigraphy and predicting geomaterial property of interest from sparse preliminary borehole data taken at a multilayered site. The proposed procedure quantifies geological and property uncertainties and implements a site investigation plan to reduce these uncertainties. The methodology involves random fields to obtain multiple simulations based upon universal kriging (UK) to predict the geomaterial property and multinomial categorical prediction (MCP) based upon spatial Markov Chains (spMC) to predict the geomaterial type. The predictions and the associated uncertainties, as quantified in terms of standard deviation (SD), are then used in determining an OSIP. The study involves several key contributions (1) Three-dimensional (3D) analysis of sparse actual borehole data is utilized to predict the project subsurface condition (2) Geological uncertainty is incorporated into the analysis through refined geomaterial layer boundaries such that all resulting layers can be analyzed with different geomaterial means. By combining the layers this way, it is possible to gain important information on the spatial structure of the whole subsurface while also improving geomaterial property prediction at the unsampled locations of the random field. The impact of incorporating geological uncertainty is examined by comparing results to when its incorporation is ignored (3) The OSIP is determined to reduce geological and property uncertainties through the geostatistical analysis of pre-existing borehole data from a bridge site. In the absence of pre-existing borehole data, borehole data are advised to be collected within the proximity of the proposed pile foundation (Goldsworthy et al. 2007; Arsyad et al. 2010). Subsequent borehole locations are selected to reduce the SD of the predicted geomaterial property. The process is repeated until the reduction effect is minimal and not worth the cost of the additional borehole. The concept behind OSIP is that of the multiphase site investigation, whereby the initial boreholes are drilled in the preliminary site investigation (preferably at the proposed foundation footprint) and the recommended boreholes are drilled in the subsequent detailed site investigation.

The paper is organized as follows. Section 2 presents the techniques and methods used in the implementation of geological uncertainty and property uncertainty to determine OSIP for bridge design. Section 3 presents the background of a case study. Section 4 compares the results with that of a recent study and validates the uncertainty simulation and quantification. The results and discussions from the analyses of the case study are presented in Sect. 5. Conclusions are given in Sect. 6.

2 Methodology

It is typically impossible to directly ascertain the geotechnical properties and the geomaterial types throughout the whole study site based on limited borehole data. However, a random field can be utilized to predict geomaterial types and properties at unobserved locations. The uncertainties associated with these predictions can be reduced through the determination of OSIP. Hence, a plan (which consists of the number and location of boreholes) is considered optimal if it minimizes uncertainties with a reasonable site investigation cost to achieve an efficient bridge design.

2.1 Geological Uncertainty

The process of analyzing geological uncertainty follows that of Oluwatuyi et al. (2023b). For a discretized grid on a random field, consider $Z(\boldsymbol{s})$ as the categorical random variable for a geomaterial type at a spatial location $\boldsymbol{s}$. Consider another spatial location $\boldsymbol{s}+\boldsymbol{h}$ where $\boldsymbol{h}$ is the multidimensional lag and $\|\boldsymbol{h}\|$ is the Euclidean distance between $\boldsymbol{s}$ and $\boldsymbol{s}+\boldsymbol{h}$ (Sartore et al. 2016). The transition probability $t_{ij}(\boldsymbol{h})$ of going from geomaterial type z_i to another z_j is.

$$t_{ij}(\boldsymbol{h}) = \Pr\left(Z(\boldsymbol{s}+\boldsymbol{h}) = z_j \middle| Z(\boldsymbol{s}) = z_i\right), \tag{1}$$

where $i,j = 1,2,\ldots,n_t$, and n_t is the number of geomaterial types (layers). The collection of transition probabilities $t_{ij}(\boldsymbol{h})$ in Eq. (1) across all i,j is given by the $n_t \times n_t$ transition probability matrix, $\boldsymbol{T}(\boldsymbol{h})$ as

$$\boldsymbol{T}(\boldsymbol{h}) = \exp\left(\|\boldsymbol{h}\|\boldsymbol{R}_{\boldsymbol{h}}\right), \tag{2}$$

where $\boldsymbol{R}_{\boldsymbol{h}}$ is the transition rate matrix which depends on the direction given by the lag $\boldsymbol{h}$. The random process of the continuous-time Markov chain will change from one geomaterial type to another according to the exponential function and as specified by the probabilities in the transition matrix. Carle and Fogg (1997) express the elements $\boldsymbol{R}_{\boldsymbol{e}_k}$ of the transition rate matrix $\boldsymbol{R}_{\boldsymbol{h}}$ in Eq. (2) as

$$\boldsymbol{R}_{\boldsymbol{e}_k} = \text{diag}\left(\{\overline{L}_{i,\boldsymbol{e}_k}\}\right)^{-1}\left(\boldsymbol{F}_{\boldsymbol{e}_k} - \boldsymbol{I}\right), \tag{3}$$

where $k = 1, 2, \ldots, d$ for d dimensions, $\boldsymbol{e}_k$ is the standard basis vector for the direction indexed by k, $\overline{L}_{i,\boldsymbol{e}_k}$ is the mean stratum thicknesses of the geomaterial type i along the direction $\boldsymbol{e}_k$, diag denotes a diagonal matrix with entries $\overline{L}_{i,\boldsymbol{e}_k}$, $\boldsymbol{F}_{\boldsymbol{e}_k}$ is the transition probability matrix consisting of probabilities for consecutive grids with the same geomaterial type along the direction $\boldsymbol{e}_k$, and $\boldsymbol{I}$ is the identity matrix The transition rate ($\boldsymbol{r}_{ij,\boldsymbol{h}}$) in row i and column j of the transition rate matrix $\boldsymbol{R}_{\boldsymbol{h}}$ in Eq. (2) is calculated as

$$\left|\boldsymbol{r}_{ij,\boldsymbol{h}}\right| = \left[\sum_{k=1}^{d}\left(\frac{h_k}{\|\boldsymbol{h}\|}\boldsymbol{r}_{ij,\boldsymbol{e}_k}\right)^2\right]^{1/2}, \tag{4}$$

where $\boldsymbol{r}_{ij,\boldsymbol{e}_k}$ denotes row i and column j of $\boldsymbol{R}_{\boldsymbol{e}_k}$ in Eq. (3), $\boldsymbol{r}_{ij,\boldsymbol{h}}$ is non-positive when $i = j$, and $\boldsymbol{r}_{ij,\boldsymbol{h}}$ is non-negative when $i \neq j$.

To predict a geomaterial type at an unknown spatial location ($\boldsymbol{s}_0$), an approximation of the conditional probability using multinomial categorical prediction in Eq. (5) is approximated as

$$q_j(\boldsymbol{s}_0) \approx \frac{p_j \prod_{l=1}^{n} t_{jk_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)}{\sum_{i=1}^{n_s} p_i \prod_{l=1}^{n} t_{ik_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)}, \tag{5}$$

where p_j is the proportion in category j and $t_{ik_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)$ represents the transition probability as defined in Eq. (1) from geomaterial type z_i to another z_j for observation with index l. Monte Carlo simulation (MCS) is thereafter applied to the predicted subsurface to propagate the uncertainty in the predictions. The conditional probabilities over the multiple simulations at each spatial location are then converted to information entropy to quantify the geological uncertainty (Wellmann and Regenauer-Lieb 2012). Consider a randomized simulation $H_m(\boldsymbol{s}_0)$ calculated at a spatial location $\boldsymbol{s}_0$ using an approximation of the conditional probability $q_{j,m}(\boldsymbol{s}_0)$ in Eq. (5) for one simulation indexed by m out of the multiple simulations n_M. The corresponding equation is given by

$$H_m(\boldsymbol{s}_0) = -\sum_{j=1}^{n_g} q_{j,m}(\boldsymbol{s}_0)\log_{n_g} q_{j,m}(\boldsymbol{s}_0) \tag{6}$$

The location average information entropy, $\overline{H}(\boldsymbol{s}_0)$ is obtained at a spatial location by taking the mean of the information entropy values in Eq. (6) summed across n_g which is the number of different geomaterial types. The logarithm is taken at a base equal to the number of the different geomaterial types. From the quantification of geological uncertainty, the stratigraphic boundaries are evaluated as continuous spatial positions with high values of mean information entropy.

2.2 Property Uncertainty

The process of analyzing geomaterial property uncertainty follows that of Oluwatuyi et al. (2022a). However, in this study, that process is carried out for each simulation of the geomaterial layer boundaries described in Sect. 2.1. By simulating the geomaterial properties using the proposed methodology, geological and property uncertainties can be analyzed collectively.

As in Oluwatuyi et al. (2022a), consider the random vector for a geomaterial property ($\mathbf{Y}$) where the covariance matrix of that geomaterial property is $\mathbb{C}(\mathbf{Y}) = \boldsymbol{\Sigma}$. The covariance matrix $\boldsymbol{\Sigma}$ is important to account for the spatial correlation structure. The spatial correlation function is specified via $\boldsymbol{H}(\rho)$ as

$$\boldsymbol{\Sigma} = \sigma^2\boldsymbol{H}(\rho) \tag{7}$$

where ρ represents the range parameter and σ^2 represents the partial sill. The exponential correlation function for $H(\rho)$ is used in this study because of its smaller information criteria value, simplicity, and application in a previous study (Oluwatuyi et al. 2022a). The exponential correlation function is

$$\boldsymbol{H}(\rho) = \exp(-h/\rho) \tag{8}$$

To predict the geomaterial property at an unobserved spatial location ($\boldsymbol{s}_0$), the best linear unbiased prediction for universal kriging is given by

$$\hat{y}(s_0) = \boldsymbol{x}(s_0)'\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{v}}'\hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{y}(\boldsymbol{s}) - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right), \quad (9)$$

where $\boldsymbol{s}$ contains the observed spatial locations, $\hat{\boldsymbol{\beta}}$ is the estimate of the regression coefficient vector $\boldsymbol{\beta}$ obtained from generalized least squares, $\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2\boldsymbol{H}(\hat{\rho})$ is the restricted maximum likelihood estimate of the covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{x}(s_0)$ is the value of the design matrix at s_0, $\boldsymbol{y}(\boldsymbol{s})$ is the vector of the geomaterial property values at $\boldsymbol{s}$, and $\hat{\boldsymbol{v}}$ is a vector of covariances with entries $\widehat{\text{cov}}(\text{Y}(s_0), \text{Y}(s_i))$ for $i = 1, \ldots, n$ (Pebesma 2004; Bivand et al. 2008).

Conditional simulation is conducted on each of the multiple simulations predicting the geomaterial type (as described in Sect. 2.1) to account for the uncertainties associated with the plug-in estimates used to perform universal kriging. The conditional simulation produces realizations that match the data at the observed locations by working with the conditional distribution given by the data (Bivand et al. 2008). In this way, the uncertainties in both the predicted geomaterial type and the predicted geomaterial are simultaneously propagated into additional analyses involving the bridge design.

2.3 Optimal Site Investigation Plan

The algorithm to obtain OSIP is implemented in the R program (R Core Team 2021). Inputs needed for the algorithm include the 3D random field measurement, transverse-longitudinal coordinates of the boreholes, the observed geomaterial type, and the geomaterial property of interest at each subsurface depth examined. The procedure for determining the OSIP for bridge design through the combined analysis of geological and property uncertainties is summarized in Fig. 1. In general, this procedure consists of twelve steps detailed as follows:

1. Discretize the design relevant bridge domain (or random field) into a 3D grid with a computationally efficient grid scheme.
2. Set out the geomaterial types as disclosed by the boreholes to the corresponding grid position. This information is conditional for the prediction of the geomaterial type and for quantifying the geological uncertainty.
3. Based on the borehole data, estimate the 3D transition rates and probabilities using spatial Markov Chains (Fabbri et al. 2020; Oluwatuyi et al. 2022b).
4. Predict the geomaterial types at unsampled locations of the discretized bridge domain using MCP and simulate geological uncertainty multiple times using MCS to generate several realizations.
5. Describe the stratigraphic boundaries of each different realization as obtained from the multiple simulations of geological uncertainty in step 4.
6. Set out the geomaterial property as disclosed by the boreholes to the described geomaterial stratification of the discretized bridge.
7. Log transform the geomaterial property and fit the spatial model with spatial correlation structure estimated based on the borehole data.
8. Predict using the UK and conditionally simulate the geomaterial property values at unsampled locations of the discretized and geomaterial type stratified bridge domain.
9. Repeat steps 6–8 for the stratigraphic boundaries of each different realization generated in step 4.
10. The sample mean and SD for each geomaterial type or layer depth are summarized across all conditional simulations and sampling locations of the four actual boreholes shown in the x–y axis on the left portion of Fig. 2. The SD is then used to assess the combined geological and property uncertainties for a given site investigation plan.
11. As shown in the heatmap on the right portion of Fig. 2, the resultant uncertainty of adding a subsequent new borehole to the actual boreholes at every possible location on the subsurface is evaluated. The recommended subsequent borehole is selected as the x–y location on the subsurface grid or heatmap with the lowest SD.
12. The actual borehole data is updated with data from the recommended borehole, and steps 10 and 11 are repeated until the SD is reduced to a point where further reduction is no longer cost-effective. The selection of the optimal number of boreholes can be visualized with a scree plot where SD (uncertainty) is plotted against the number of boreholes (cost) as shown in Fig. 3.

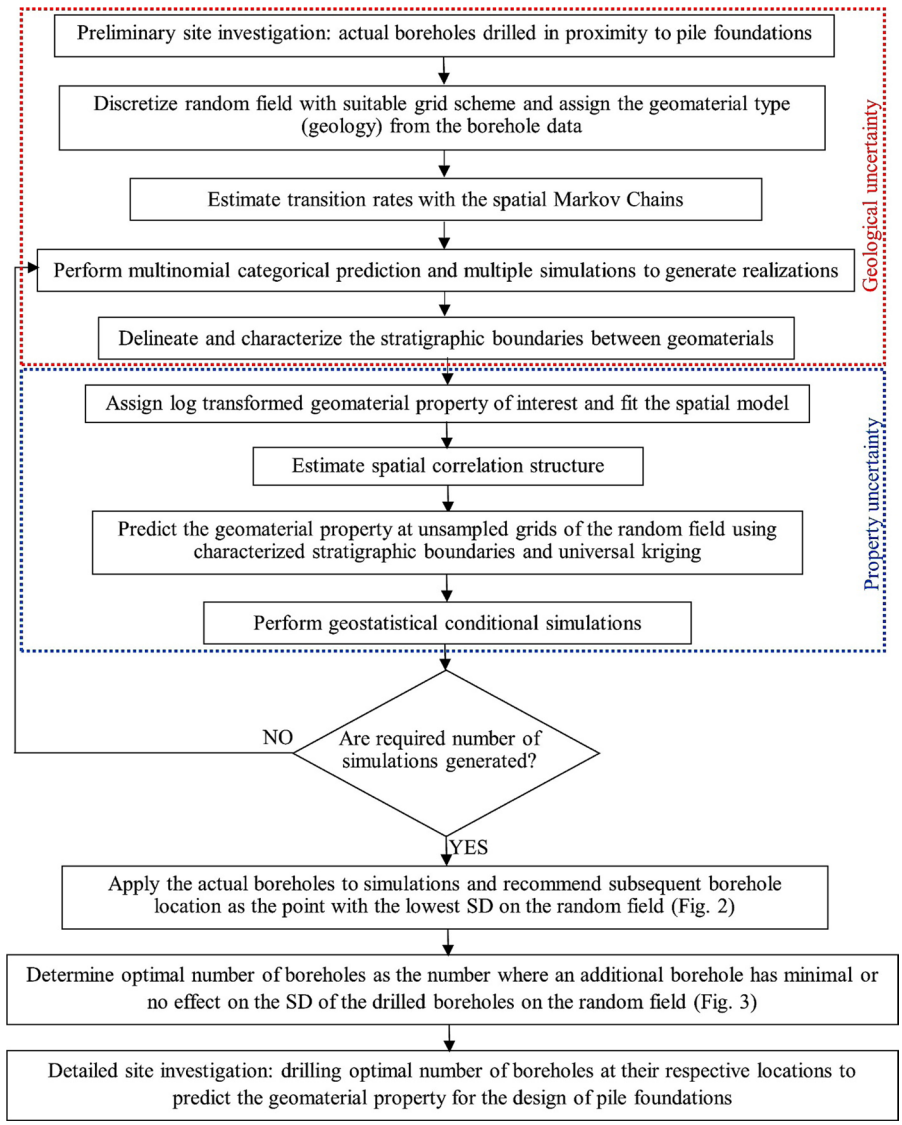

Fig. 1 Flowchart of an optimal site investigation plan for bridge design considering both geological and property uncertainties

The computational time for OSIP depends on several factors including the study site area, the use of 3D analysis, discretized grid size selected, the number of multiple simulations selected to propagate the uncertainties, and the computational power of the computer system. For piles bearing on soft rocks or intermediate geomaterials (IGMs), a minimum of 4.5 m depth of continuous bedrock or IGM core is recommended. Alternatively, the borehole is terminated at a maximum depth of 30 m (AASHTO 2020). Discrete in-situ tests, like the standard penetration test (SPT), and sampling for laboratory tests are recommended to be taken at every 1.5 m depth (Oluwatuyi et al. 2022a). OSIP determination in this study is not directly related to the response of the bridge pile foundation. This is because the degree of uncertainty in the geomaterial properties is higher and difficult to characterize compared to that of the structural component. Moreover, details of the structural component, such as pile size and the load imposed, may not be available at the time of planning and executing the subsurface investigation. This is in line with current engineering practice where the subsurface investigation is usually done before geotechnical and structural design. However, with the minimization of geological and property uncertainties through OSIP, a more accurate prediction of the geomaterial property can be obtained. With a more accurate prediction and

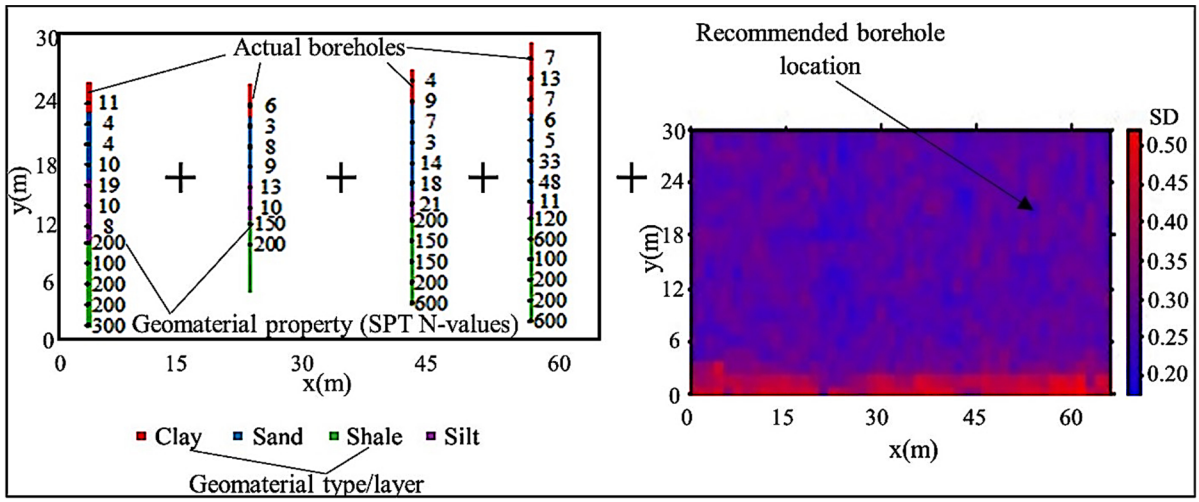

Fig. 2 Pictorial explanation of subsequent recommended borehole selection for OSIP (plus sign implies combination and not summation)

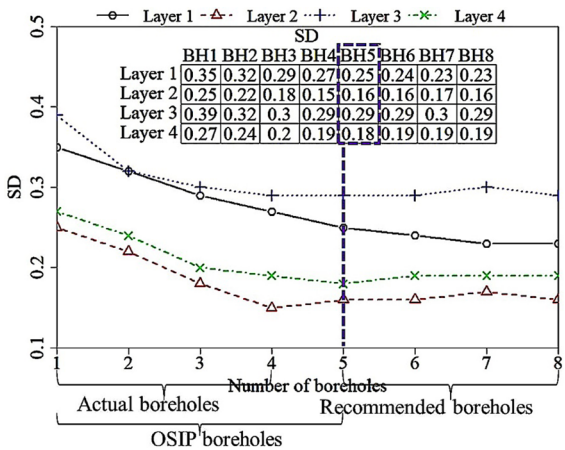

Fig. 3 Pictorial explanation of actual and recommended boreholes that determine the optimal number of boreholes (Ng et al. 2023)

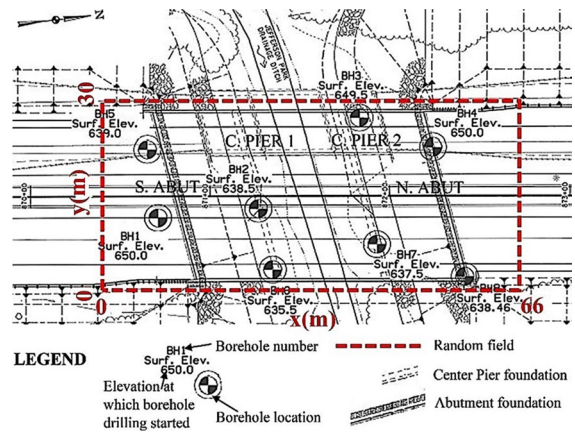

Fig. 4 Eight boreholes at the bridge site in Wapello County, Iowa (Ng et al. 2023)

quantifiable subsurface uncertainties, the proposed OSIP procedure can be incorporated into the framework of the load and resistance factor design for the bridge foundation (Oluwatuyi et al. 2023a).

3 Case Study

The data for this study is from the SPT conducted on the project site by the Iowa Department of Transportation (IADOT) to replace an existing US 63 bridge and culverts over a drainage ditch about 0.3 miles south of Ottumwa, Wapello County, Iowa. Eight boreholes are drilled on the project site as shown in Fig. 4. SPT is a common in-situ test used by engineers to obtain geomaterial samples from the boreholes and its N-value is an indirect measurement of geomaterial resistance. The N-value is the number of hammer blows per m by which a 63.5 kg hammer with a 0.75 m drop height can drive a split spoon sampler 0.3 m into the geomaterial. For the shale (IGM) layer, a full 0.3 m penetration is not achieved

by the sampler. Hence, measured blows are linearly extrapolated using the procedure described in Oluwatuyi et al. (2022a) for comparable and consistent analysis. In Sect. 5 of this paper, a total of 46 SPT N-values from the four boreholes in proximity to the four bridge foundation locations are used as shown in Fig. 5. The foundation locations on the study site are namely north abutment (N. ABUT.), center pier 1 (C. PIER. 1), center pier 2 (C. PIER. 2), and south abutment (S. ABUT.). The four remaining boreholes expunged from the OSIP analysis are used as validating data. Also, their relative locations are compared with the subsequently recommended boreholes from the OSIP analysis. The subsurface is characterized according to the borehole data into four geomaterial layers with layer 1 consisting of dark brown-gray stiff clay (denoted as Clay), layer 2 consisting of brown-gray clayey/fine sand to coarse sand, (denoted as Sand), layer 3 consisting of the gray firm to stiff silt with brown firm sand clay with brown-gray firm clay and gravel (denoted as Silt), and layer 4 as the IGM layer consisting of gray slightly to moderately weathered shale (Shale).

The idea behind the method for OSIP proposed in this study is to take the borehole data from the preliminary site investigation and analyze it for spatial positions that will reduce uncertainty for all layers or a target layer. The actual borehole data from the preliminary site investigation is recommended to be taken

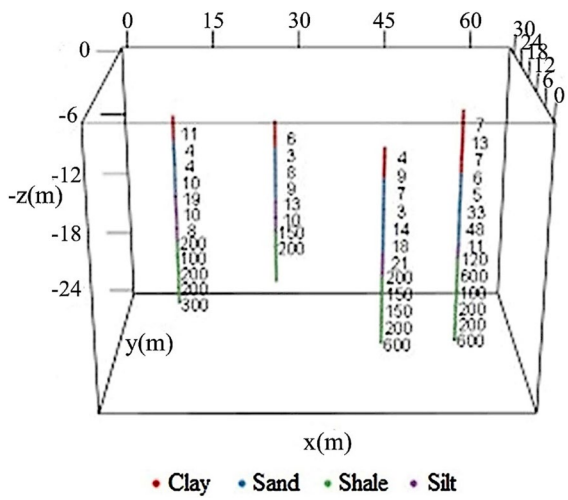

Fig. 5. 46 SPT N-values in four geomaterial layers as observed from the four drilled boreholes in proximity to the foundation locations of the study site (Ng et al. 2023)

in proximity to the proposed foundation to reduce the uncertainties that may be propagated to the design. However, it should be noted that the preliminary site investigation of locating boreholes in the proximity of the foundation is not sufficient for a design as there are still unknowns (uncertainties) within the site. Also, piles are designed and constructed in groups, and borehole information can at best be limited to the design of a pile closest to the borehole. The 3D random field for the study site has dimensions + 66 m by + 30 m by − 24 m for the x, y, and z axes, respectively. The negative sign in the z-axis indicates depth below the ground surface. The SPT N-values are log transformed as is typical with geomaterial properties due to their wide range of values (Grasmick et al. 2020; Oluwatuyi et al. 2022a).

4 Model Comparison and Validation

4.1 Model Comparison

The model considering both geological and property uncertainties in determining an OSIP in Sect. 2 of this present study is compared to the related model where only property uncertainty is considered. Twelve boreholes comprised of eight actual preliminary boreholes (Fig. 4) and four recommended boreholes initially determined for uncertainty reduction for a bridge project in Wapello County Iowa are analyzed for their uncertainties using the model in the present study. The uncertainties in terms of SDs present in each layer are compared for the conditional simulation as shown in Fig. 6. The term 'discrepancies', denoted as DISC, in Fig. 6 is a measure of the difference in the quantified uncertainties between the present study considering both geological and property uncertainties denoted as PUGU and one considering only property uncertainty denoted as PU. The uncertainty discrepancies in both models for layers 1, 2, and 4 as earlier described in Sect. 3 are within ± 0.1 with an increase in the number of boreholes. However, for layer 3, the uncertainty discrepancy is as high as 0.3 at three boreholes. These high discrepancies in the quantified uncertainties from both models for layer 3, which is the transitional layer before the Shale layer (layer 4), may be due to the thinness of the layer as revealed by the boreholes in Fig. 5. Interestingly, for layer 3, the uncertainty discrepancy is as

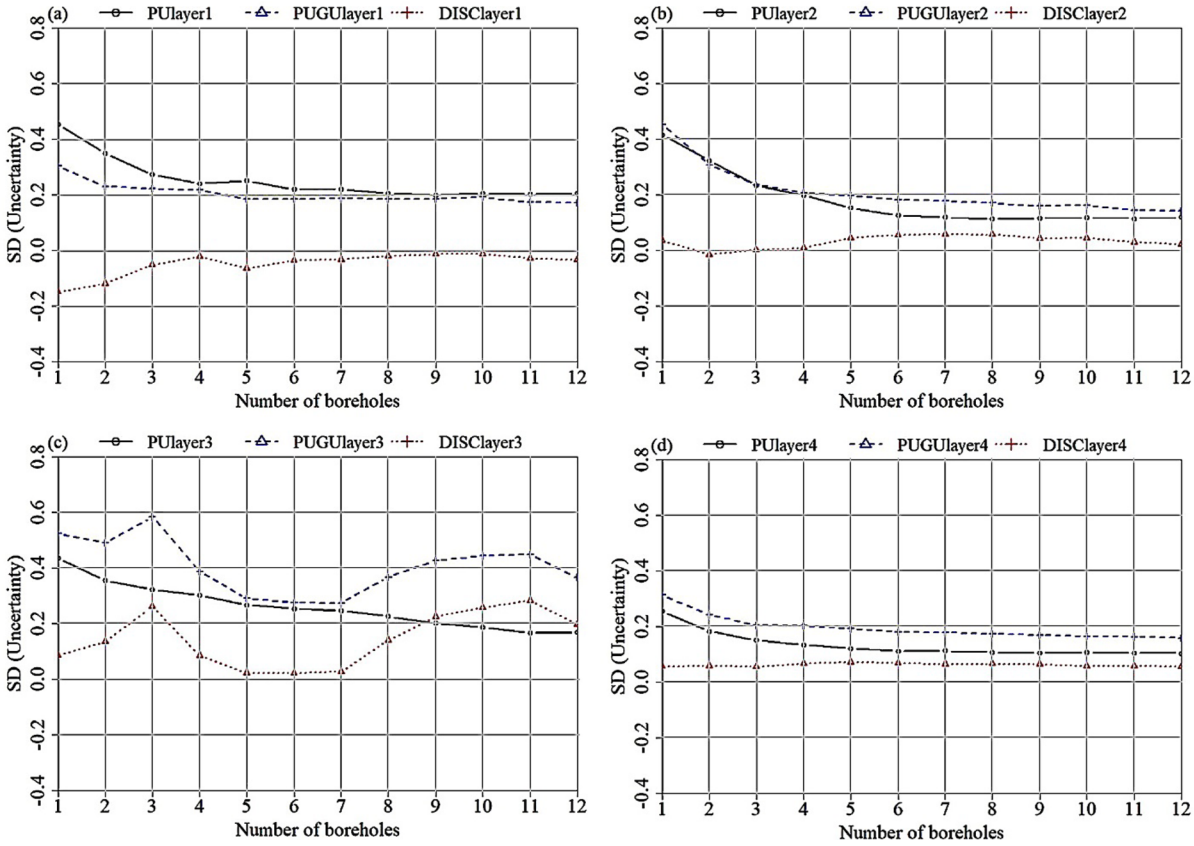

Fig. 6 PU and PUGU model comparison in the different geomaterial layers (Ng et al. 2023)

low as 0.02 at six boreholes which is the same as the optimal number of boreholes recommended for the study considering property uncertainty alone (Oluwatuyi et al. 2022a). The true effect of incorporating geological uncertainty may not be truly known unless predictions from both models are validated.

4.2 Model Validation

Validation of a model can provide important insights into the practical viability of resulting predictions. The geomaterial property from observed borehole data $\left(y_{ij}\right)$ is used to assess the model prediction accuracy since these observations give the measured value of the geomaterial property response. Here, $j = 1, 2, \ldots, b$ indexes b boreholes and $i = 1, \ldots, n_b$ indexes n_b observations within a borehole. Leave one-out cross validation (LOOCV) is used because it will be performed on the entire data resulting in the estimation of the same test error. LOOCV is also important in this study since there is no comparable test data available for the site. For the application of LOOCV, the data is first split into two parts where one part is the training data which consists of all the observed boreholes, except for the one borehole used in the second part for testing data. The training data is used to predict the geotechnical property for the n_b observations within the testing borehole $\left(\hat{y}_{(ij)}\right)$ and compared to the observed or measured geotechnical property for the testing borehole $\left(y_{(ij)}\right)$. The process is then repeated for each of the b available boreholes. The combined evaluation of these predictions involving the testing boreholes is obtained by the root mean squared error (*RMSE*) in Eq. 10. The predicted geomaterial property on the subsurface is used because it is more suitable to assess model accuracy through it rather than a smoothened mean estimate from multiple simulations (Goovaerts 2001). Bias is also calculated for the geomaterial property prediction

in each observation within a borehole. The bias for each test is defined by Eq. 11.

$$RMSE = \frac{1}{b}\sum_{j=1}^{b}\left[\frac{1}{n_b}\sum_{i=1}^{n_b}\left(\hat{y}_{(ij)} - y_{(ij)}\right)^2\right]^{\frac{1}{2}} \quad (10)$$

$$Bias_{(ij)} = \frac{y_{(ij)}}{\hat{y}_{(ij)}} \quad (11)$$

The model accuracy plot with the bias distribution of the PUGU and PU models and the corresponding RMSE is shown in Fig. 7. The bias distribution plot shows that the PUGU model had its mean bias value equal to 1. The coefficient of variation (COV) of the bias in the PUGU model is 0.09 compared to 0.14 in the PU model. The trend is also consistent with the RMSE obtained from the cross validation where the PUGU model has a lower mean prediction error of 0.18 compared to the 0.34 obtained from the PU model. In other words, consideration of geological uncertainty in the prediction of geomaterial property decreases the prediction error on an average from 34 to 18%. Worthy of note is the high RMSE value recorded when borehole 6 is used as the testing data. Recall, that six boreholes are recommended as the optimal number of boreholes by Oluwatuyi et al. (2022a).

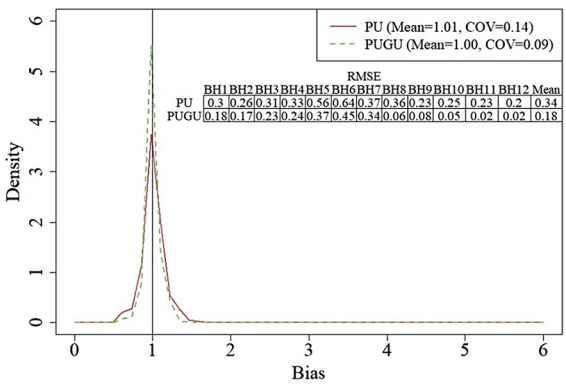

Fig. 7 Accuracy plot showing the bias distribution and RMSE values for cross validation of the PU and PUGU models (Ng et al. 2023)

5 Results and Discussion

5.1 Geomaterial Type Prediction

The estimate of the transition rate matrices in Eq. 3 using the borehole data is presented in Table 1. From these transition rates, the transition probabilities are estimated using Eq. 2. A transiogram model is fit to the estimated transition probabilities in the Z-direction (depth) as shown in Fig. 8. The transiogram is analogous to the variogram used in the prediction of a geomaterial property at unobserved spatial locations. The transiogram represents the conditional probabilities of an unobserved geomaterial along a certain directional axis and Euclidean distance (lag) from an observed geomaterial at any spatial point. For instance, for the transiogram in the z-direction at lag 50 (Fig. 8), there is a 75% probability that Shale will transition to Shale, a 10% probability that Shale will transition to Sand, 8% probability that Shale will transition to Clay and 7% probability that Shale will transition to Silt. From Table 1, we can also infer that the transition rates in the Y-direction and Z-direction are higher than in the X-direction which may be due to the lengths and the layout of the geomaterials. The negative values in the diagonal of the transition rate matrices in Table 1 are consistent with the inverse relationship

Table 1 3D estimated transition rate matrices obtained using the maximum entropy method (Ng et al. 2023)

Geomaterial class	Clay	Sand	Silt	Shale
X-direction estimated transition rate matrix				
Clay	− 0.012	0.005	0.003	0.004
Sand	0.003	− 0.009	0.003	0.003
Silt	0.003	0.004	− 0.010	0.003
Shale	0.002	0.003	0.002	− 0.007
Y-direction estimated transition rate matrix				
Clay	− 0.055	0.020	0.013	0.022
Sand	0.016	− 0.055	0.015	0.024
Silt	0.014	0.020	− 0.055	0.021
Shale	0.017	0.023	0.015	− 0.055
Z-direction estimated transition rate matrix				
Clay	− 0.039	0.013	0.013	0.013
Sand	0.012	− 0.054	0.021	0.021
Silt	0.026	0.044	− 0.114	0.044
Shale	0.011	0.019	0.019	− 0.049

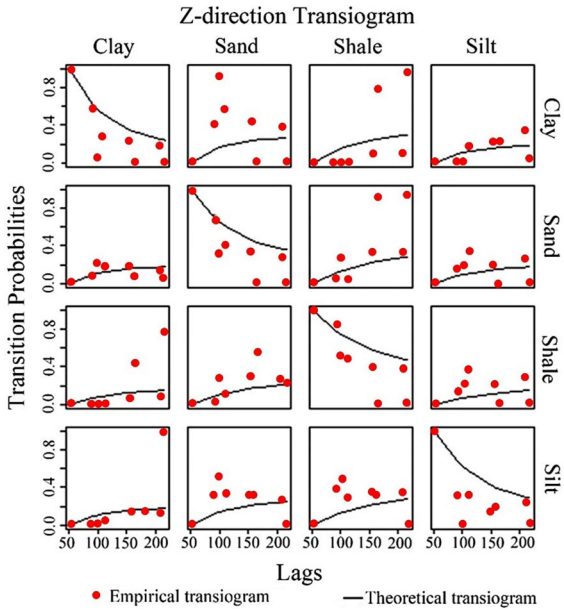

Fig. 8 Z-direction transiogram estimated from the data (Ng et al. 2023)

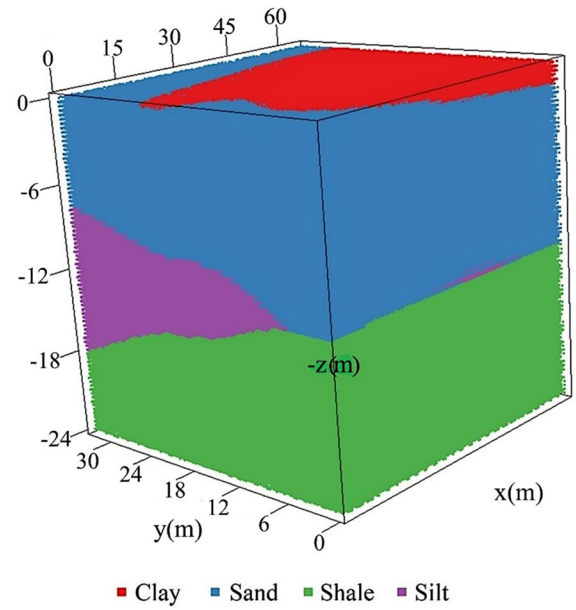

Fig. 9 Simulated subsurface showing predicted geomaterial type (Ng et al. 2023)

of the graph in the diagonal of the transiogram in Fig. 8

The simulated subsurface averaged over multiple simulations for the geomaterial type is shown in Fig. 9. A total of 100 simulations are deemed adequate as additional simulations resulted in less than ± 1% change in the information entropy (Pyrcz and Deutsch 2014). Oluwatuyi et al. (2022b) and Oluwatuyi et al. (2023b) have also shown this number of simulations to be adequate, as additional simulations do not appear to justify the additional computation time. This simulated geomaterial type in the random field contains the geological uncertainties to be considered along with property uncertainties. The incorporation of geological uncertainty will enhance the separation of the four geomaterial types on the simulated subsurface in Fig. 9 into layers through highly nonlinear spatial boundaries, unlike previous methods of treating the subsurface as a single layer or multiple layers determined through the linear interpolation of the historical geological data. These spatially determined layers are thereafter analyzed collectively with different averages of the geomaterial property by layer to accurately predict the subsurface geomaterial property.

5.2 Geomaterial Property Prediction

The SPT N-values measured from the boreholes are log transformed and the average for each of the four spatially determined layers is estimated. Allowing for different means across the layers accounts for geomaterial heterogeneity and possible violation of weak stationarity. This approach also allows for homogeneous spatial correlation to be modeled across the study site which is especially critical for modeling sparse borehole data. The exponential correlation structure is used because it incorporates spatial correlation that is similarly used by Oluwatuyi et al. (2022a). Assumptions of anisotropy and isotropy are checked by fitting these models for log SPT N-values with the Exponential spatial structure in the X–Z and Y–Z directions and by comparing the information criteria presented in Table 2. The isotropic model is preferable for both combinations based upon smaller AIC and BIC. This implies that the covariance structure does not change with direction (Bivand et al. 2008). A variogram of log SPT N-values using the fit of an exponential model and no nugget is shown in Fig. 10. Prediction of log SPT N-values for the subsurface is done using universal kriging such that the different layers are analyzed with the respective average of log

Table 2 Information criteria (smaller-is-better) of the GLM with Exponential spatial correlation for the X–Z and Y–Z coordinate systems (Ng et al. 2023)

GLM	X–Z				Y–Z			
	n2llik	AIC	AICc	BIC	n2llik	AIC	AICc	BIC
Isotropic	97.08	109.08	111.24	119.51	95.43	107.43	109.58	117.85
Anisotropic	96.97	112.97	116.87	126.87	92.92	108.92	112.81	122.82

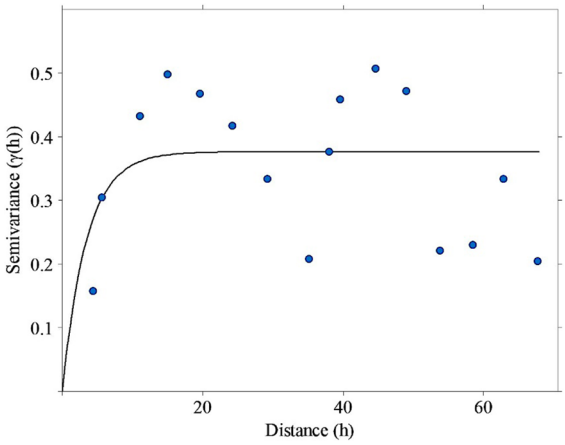

Fig. 10 Isotropic variogram of the exponential spatial correlation for SPT N-values (Ng et al. 2023)

SPT N-values. Log SPT N-values are conditionally simulated for each of the 100 four-layered subsurfaces. The log SPT N-values simulated subsurface is back-transformed to obtain SPT N-values in blows/m on the original scale. The simulated subsurfaces are averaged for the log and back-transformed predicted SPT N-values as shown in Fig. 11. The determination of the recommended borehole locations and the borehole numbers for the OSIP is based on the 100 conditionally simulated subsurfaces containing the log predicted SPT N-values.

5.3 Optimal Site Investigation plan (OSIP)

Site investigation plans with relative borehole drilling locations obtained from IADOT and conditional simulations are compared in Fig. 12. The borehole locations from the IADOT sampling plan are selected based on current site investigation practice. The boreholes from the conditional simulations are drilled at locations where both geological and property uncertainties are minimized for the shale or IGM layer over the 100 conditional simulations of predicted log SPT N-values. Uncertainty minimization is specified

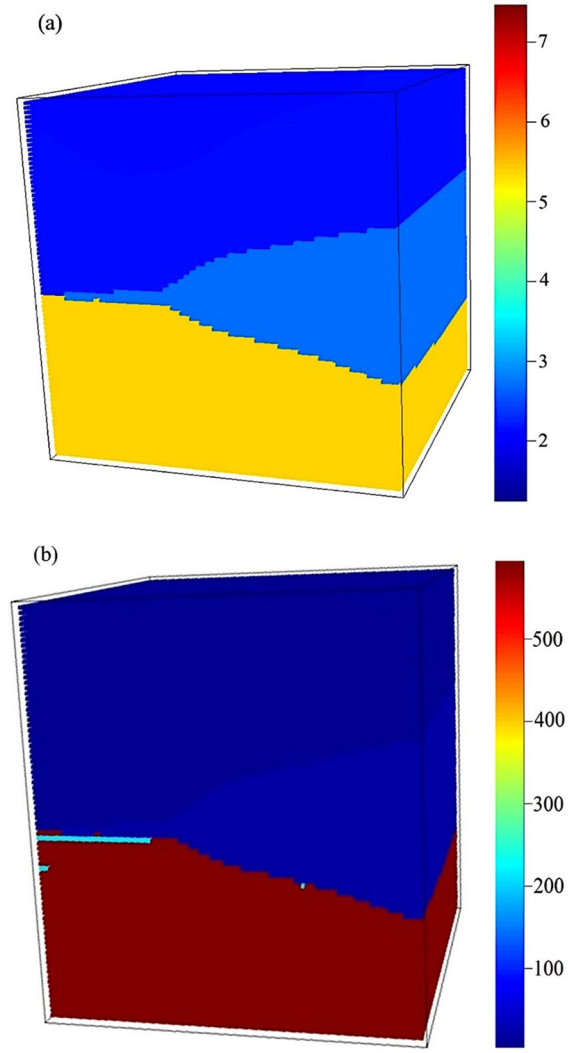

Fig. 11 Simulated subsurface showing **a** log and **b** back-transformed predicted SPT N-values (Ng et al. 2023)

for the shale layer because most of the piles used in the bridge construction at the case study site are end-bearing piles. The measure of uncertainty in this study by SD on the logarithmic scale is approximately equal to the coefficient of variation (COV) on the arithmetic scale (Julious and Debarnot 2000).

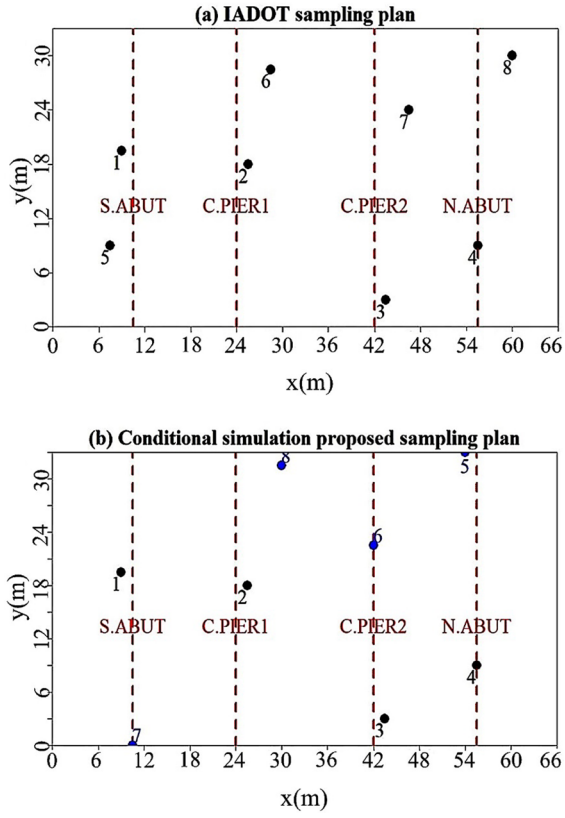

Fig. 12 Relative borehole drilling locations for **a** IADOT sampling plan and **b** conditional simulation proposed sampling plan (Ng et al. 2023)

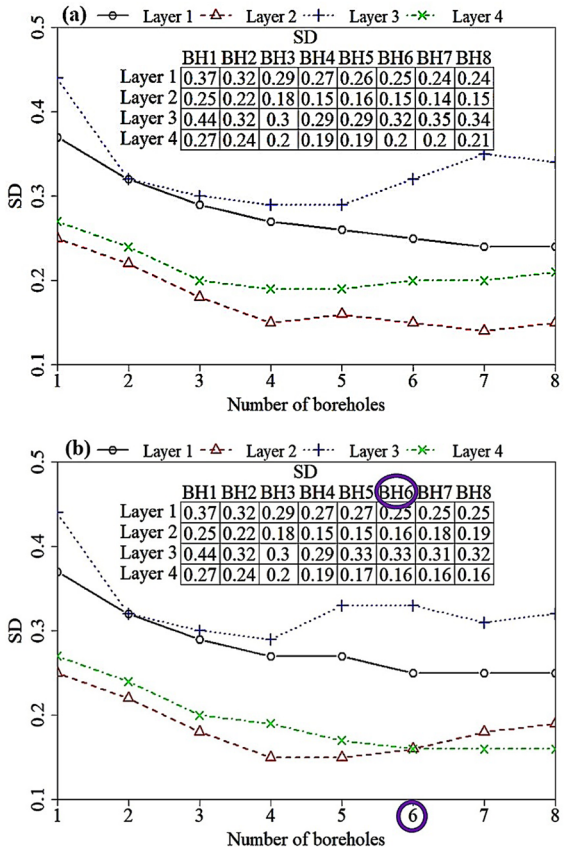

Fig. 13 Uncertainties in the four geomaterial layers as obtained from **a** IADOT sampling plan and **b** conditional simulation proposed sampling plan (Ng et al. 2023)

Boreholes 1 to 4 which are in the proximity of the four foundation locations in Fig. 12 provide the actual or preliminary borehole information from the case study. The relative locations for the subsequent boreholes 5 to 8 differ for the two sampling plans (IADOT and conditional simulation). The conditional simulation proposed sampling plan, just like the IADOT sampling plan, has geomaterial profile information maximized with boreholes spread across the site. Uncertainties are quantified in Fig. 13 for the four geomaterial layers of the different borehole locations and two sampling plans on the multiple simulated random fields. The idea is to quantify the uncertainties due to the actual boreholes numbered 1 to 4 and determine how the relative locations of subsequent boreholes numbered 5 to 8 will reduce these uncertainties. The optimal number is selected as the number of boreholes with the minimum SD in the shale as the pile end-bearing layer in which further reduction in SD is no longer worth the additional effort. For the conditional sampling plan, the optimum number of boreholes is 6 as shown in Fig. 12b. The conditional simulation proposed plan has the lowest SD of 0.155 at 6 boreholes compared to the lowest SD of 0.187 obtained from the IADOT sampling plan at 4 boreholes. It is reasonable to compare four boreholes to six boreholes as will be done in Sect. 5.4, because we are comparing the best performance of these site investigation plans which are at their lowest uncertainties. In most bridge designs, the conditional simulation proposed sampling plan is ideal for subsurface characterization because it is site-specific. Also, the preliminary borehole information from the foundation locations will be sufficient for spatial analysis of geological and property uncertainties to recommend subsequent locations of boreholes for the detailed site investigation.

5.4 OSIP Validation

It is necessary to validate the proposed OSIP boreholes to see if it is an adequate representation of the study site. The previous validation conducted in Sect. 3 is done to determine the prediction accuracy of the model and to demonstrate the need to consider geological uncertainty. Validation in this section is done to determine the prediction accuracy of the conditional simulation and to show the need for the recommended boreholes from the OSIP. Boreholes 1 to 6 from the conditional simulation sampling plan in Fig. 12b are denoted as the OSIP boreholes (OSIP BHs). These OSIP BHs are a combination of four actual preliminary boreholes (which are also boreholes 1 to 4 in the IADOT sampling plan denoted as Actual BHs), and two recommended boreholes (boreholes 5 and 6) selected to reduce uncertainty. The six boreholes (OSIP BHs) are used as model training data to simulate the geomaterial types and property (predicted log SPT N-values) for the whole subsurface. The four boreholes (boreholes 5 to 8) earlier expunged in Sect. 3 from the IADOT sampling plan as shown in Figs. 4 and 12a are used as model testing data to validate the predicted geomaterial types and log SPT N-values earlier obtained. Once the whole subsurface is simulated using data from the four actual and six OSIP boreholes. The predicted geomaterial types in all four layers of the simulated subsurface at the exact locations of boreholes 5 to 8 are compared to the geomaterial types observed in all four layers of boreholes 5 to 8 to determine their percentage match (prediction accuracy) as shown in Fig. 14. The percentage match results as summarized in Table 3 show that the six OSIP boreholes have an average 67.4% geomaterial type prediction accuracy compared to the average 62.7% geomaterial type prediction accuracy estimated when four actual boreholes are used as model training data. The rationale for comparing four actual boreholes to

Table 3 Geomaterial type prediction accuracy (Ng et al. 2023)

Site investigation plan	Percentage match for boreholes 5 to 8 in the IADOT sampling plan				
	BH5	BH6	BH7	BH8	Mean
4 Actual Boreholes	62.9	54.8	68.1	64.9	62.7
6 OSIP Boreholes	68.0	57.8	70.0	73.8	67.4

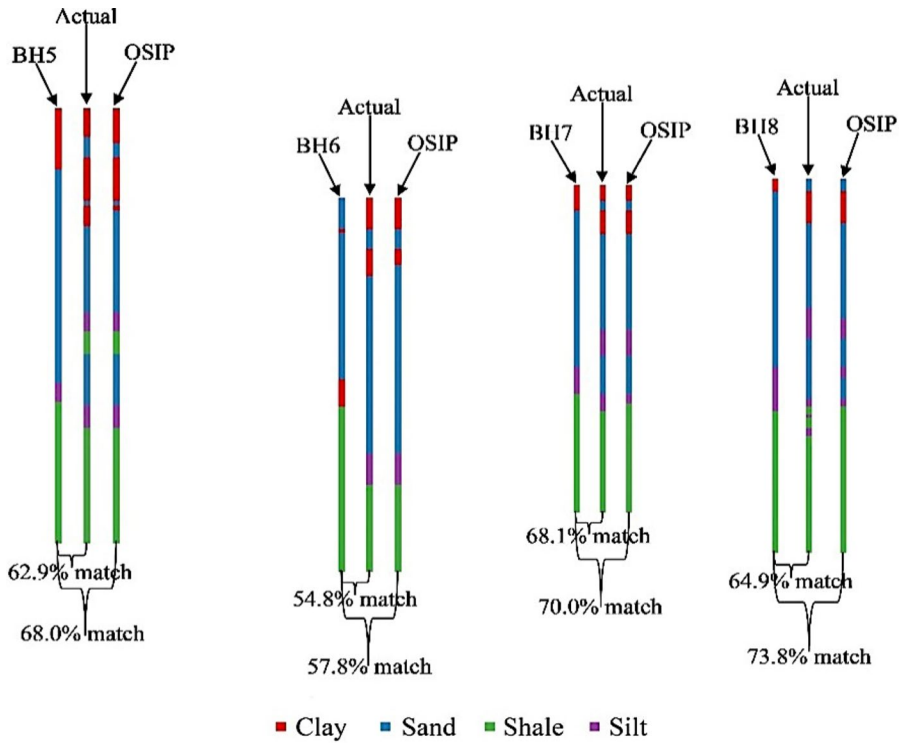

Fig. 14 Comparison between the observed borehole geomaterial types and the predicted geomaterial types from the actual and OSIP boreholes (Ng et al. 2023)

the six OSIP boreholes had been described earlier in Sect. 5.3 of the paper.

Predicted log SPT N-values at the exact locations of boreholes 5 to 8 on the simulated subsurface are also compared to the measured log SPT N-values to assess prediction accuracy. The results of validating the geomaterial property using the data from the actual preliminary boreholes and the OSIP boreholes are shown in Fig. 15. The results also show that the OSIP boreholes have a better prediction accuracy than the actual preliminary boreholes. The prediction error measured in terms of RMSE (Eq. 11) decreased from 27 to 10% with the proposed OSIP. Figure 15 shows that the validation of the OSIP boreholes is much closer to the equality line compared to that of the actual boreholes with a mean bias of 1.01 closer to unity and a smaller SD of 0.04. This validates the need for the recommended boreholes through the proposed OSIP resulting in a more accurate and consistent prediction of the geomaterial property of the subsurface. In particular, the improved prediction and reduction in uncertainties of a critical geomaterial layer, such as the shale in this case study, invariably improve the design efficiency and performance of an end-bearing pile foundation.

6 Summary and Conclusions

This study proposed a method for determining the OSIP considering both geological uncertainty and property uncertainty for bridge foundation design. Geological uncertainty is addressed to better analyze a multilayered site. Analysis of the sparse borehole data from a preliminary site investigation is conducted in 3D to accurately represent the project subsurface. In the absence of pre-existing borehole data on the site, boreholes are recommended at the proposed pile foundation footprint. The methodology involves predicting the geomaterial type and the geomaterial property at unobserved spatial positions of the random field using multinomial categorical prediction and universal kriging, respectively. Geological and property uncertainties are obtained for these predictions through multiple simulations. The method is compared to that in which only property uncertainty is considered and is also validated by assessing prediction accuracy and consistency. The following important conclusions are drawn from this study:

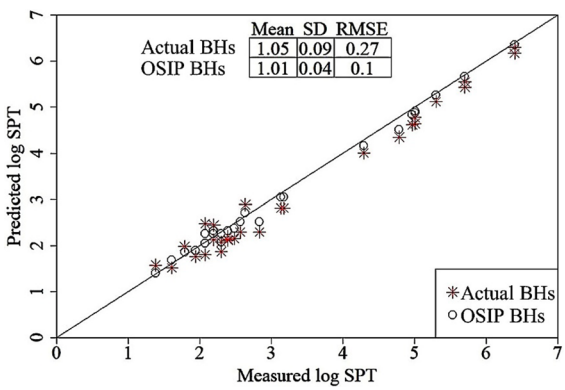

Fig. 15 Comparison of measured and predicted log SPT for the OSIP boreholes and the actual boreholes (Ng et al. 2023)

- Cross validation of models shows that the average 34% prediction error of the model considering property uncertainty alone is higher than the average 18% estimated for the model considering both geological and property uncertainties. Hence, the consideration of both geological uncertainty and property uncertainty is important to improve the model prediction accuracy.
- Compared to the IADOT plan with four boreholes and the lowest SD of 0.187, the OSIP determined from the conditional simulation for the IGM layer (layer 4) results in the lowest SD of 0.155 at six boreholes. The lower uncertainty resulting from the OSIP improves the design efficiency of bridge foundations.
- The validation for six OSIP boreholes consisted of four actual boreholes (same as boreholes 1 to 4 of the IADOT plan) and two recommended boreholes, showing a decrease in the prediction error from 27 to 10%. This finding supports the application of the proposed OSIP considering the addition of new recommended boreholes.
- It is important to consider minimizing both geological and property uncertainties in the development of OSIP. Considering both uncertainties in OSIP provides a better prediction and reduces the uncertainty of the geomaterial response for bridge pile foundations.

This paper provides the methodology and justification for developing the OSIP. While the study was based on intermediate geomaterials from the

sedimentary terrain, the geostatistical methods used in this study are also applicable to igneous and metamorphic terrains. Future studies will look at the incorporation of OSIP into the subsequent estimation of pile resistances for improving the reliability analysis and design of bridge pile foundations.

Acknowledgements The State of Wyoming, WYDOT, and the University of Wyoming reserve a royalty-free, nonexclusive, unlimited, and irrevocable license to reproduce, published, or otherwise use, and to authorize others to use the copyright in any work that is generated from the research project entitled Wyoming, Comprehensive Field Load Test, and Geotechnical Investigation Program for Development of LRFD Recommendations of Driven Piles on Intermediate GeoMaterials, TPF 5-391.

Authors' contributions All authors have reviewed and approved this manuscript.

Funding This study is supported by the Wyoming Department of Transportation as the lead agency, the Colorado Department of Transportation, the Iowa Department of Transportation, the Kansas Department of Transportation, the North Dakota Department of Transportation, the Idaho Transportation Department, and the Montana Department of Transportation under Grant RS05219.

Availability of data and material Data will be provided upon request.

Code availability Not applicable.

Declarations

Conflict of interest No conflicts of interest.

References

AASHTO (2020) Section 10 - Foundation. In: AASHTO LRFD bridge design specification. American Association of State Highway and Transportation Officials, Washington, DC

Abdulai M, Sharifzadeh M (2019) Uncertainty and reliability analysis of open pit rock slopes: a critical review of methods of analysis. Geotech Geol Eng 37(3):1223–1247. https://doi.org/10.1007/s10706-018-0680-y

Arsyad A, Jaksa M, Kaggwa W, Mitani Y (2010) Effect of Radial Distance of a Single CPT Sounding on the Probability of Over-and Under-Design of Pile Foundation. In: Proceedings of the First Makassar international conference on civil engineering (MICCE2010). Makassar, Indonesia., pp 1–5

Bivand RS, Pebesma EJ, Gómez-Rubio V, Pebesma EJ (2008) Applied spatial data analysis with R. Springer, New York

Bock H (2006) Common ground in engineering geology, soil mechanics and rock mechanics: past, present and future. Bull Eng Geol Environ 65(2):209–216. https://doi.org/10.1007/s10064-006-0059-9

Boumezerane D, Belkacemi S, Žlender B (2014) Site soundings density for geotechnical investigation with combined fuzzy and probabilistic input information. Geotech Geol Eng 32(2):547–559. https://doi.org/10.1007/s10706-014-9733-z

Carle SF, Fogg GE (1997) Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Math Geol 29(7):891–918. https://doi.org/10.1023/A:1022303706942

Crisp MP, Jaksa MB, Kuo YL (2020) Toward a generalized guideline to inform optimal site investigations for pile design. Can Geotech J 57(8):1119–1129. https://doi.org/10.1139/cgj-2019-0111

Deng ZP, Li DQ, Qi XH et al. (2017) Reliability evaluation of slope considering geological uncertainty and inherent variability of soil parameters. Comput Geotech 92:121–131. https://doi.org/10.1016/j.compgeo.2017.07.020

Fabbri P, Gaetan C, Sartore L, Libera ND (2020) Subsoil reconstruction in geostatistics beyond kriging: a case study in Veneto (NE Italy). Hydrology 7(1):1–15. https://doi.org/10.3390/hydrology7010015

Goldsworthy JS, Jaksa MB, Fenton GA et al. (2007) Effect of sample location on the reliability based design of pad foundations. Georisk Assess Manag Risk Eng Syst Geohazards 1(3):155–166. https://doi.org/10.1080/17499510701697377

Gong W, Zhao C, Juang CH et al. (2020) Stratigraphic uncertainty modelling with random field approach. Comput Geotech 125:103681. https://doi.org/10.1016/j.compgeo.2020.103681

Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103(1–2):3–26. https://doi.org/10.1016/S0016-7061(01)00067-2

Grasmick JG, Mooney MA, Trainor-Guitton WJ, Walton G (2020) Global versus local simulation of geotechnical parameters for tunneling projects. J Geotech Geoenvironmental Eng 146(7):04020048. https://doi.org/10.1061/(asce)gt.1943-5606.0002262

Han W, Yan Y, Yan X (2020) Uncertainty and sensitivity analysis of in-situ stress in deep inclined strata. Geotech Geol Eng 38(3):2699–2712. https://doi.org/10.1007/s10706-019-01179-3

Jelušič P, Žlender B (2014) An adaptive network fuzzy inference system approach for site investigation. Geotech Test J 37(3):400–411. https://doi.org/10.1520/GTJ20120018

Juang CH, Zhang J, Shen M, Hu J (2019) Probabilistic methods for unified treatment of geotechnical and geological uncertainties in a geotechnical analysis. Eng Geol 249:148–161. https://doi.org/10.1016/j.enggeo.2018.12.010

Julious SA, Debarnot CAM (2000) Why are pharmacokinetic data summarized by arithmetic means? J Biopharm Stat 10(1):55–71. https://doi.org/10.1081/BIP-100101013

Liao K, Wu Y, Miao F et al. (2022) Probabilistic risk assessment of earth dams with spatially variable soil properties using random adaptive finite element limit analysis. Eng Comput. https://doi.org/10.1007/s00366-022-01752-0

Mazraehli M, Zare S (2022) Probabilistic estimation of rock load acting on tunnels considering uncertainty in peak and post-peak strength parameters. Geotech

Geol Eng 40(5):2719–2736. https://doi.org/10.1007/s10706-022-02057-1

Mendoza C, Hurtado JE (2022) The importance of geotechnical random variability in the elastoplastic stress-strain behavior of shallow foundations considering the geological history. Geotech Geol Eng 40(7):3799–3818. https://doi.org/10.1007/s10706-022-02132-7

Ng KW, Masud NB, Oluwatuyi OE, Wulff SS (2023) Comprehensive field test and geotechnical investigation program for development of LRFD recommendations of driven piles on intermediate geomaterials. Cheyenne, WY

Oluwatuyi OE, Holt R, Rajapakshage R et al. (2022a) Inherent variability assessment from sparse property data of overburden soils and intermediate geomaterials using random field approaches. Georisk Assess Manag Risk Eng Syst Geohazards 16(4):766–781. https://doi.org/10.1080/17499518.2022.2046783

Oluwatuyi OE, Rajapakshage R, Wulff SS, Ng KW (2022b) Quantifying geological uncertainty using conditioned spatial Markov chains. Geo-Congress 2022:436–445

Oluwatuyi OE, Ng KW, Wulff SS, Masud NB (2023a) The Effect of Geological Uncertainty on the Shaft Resistance Prediction and Reliability of Piles Driven in Multi-Layered Geomaterials. Transp Res Rec. https://doi.org/10.1177/03611981221149733

Oluwatuyi OE, Rajapakshage R, Wulff SS, Ng KW (2023b) Proposed hybrid approach for three-dimensional subsurface simulation to improve boundary determination and design of optimum site investigation plan for pile foundations. Soils Found 63(1):101269. https://doi.org/10.1016/J.SANDF.2022.101269

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30(7):683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pyrcz MJ, Deutsch CV (2014) Geostatistical reservoir modeling. Oxford University Press, New York

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Sartore L, Fabbri P, Gaetan C (2016) spMC: an R-package for 3D lithological reconstructions based on spatial Markov chains. Comput Geosci 94:40–47. https://doi.org/10.1016/j.cageo.2016.06.001

Shi C, Wang Y (2021) Smart determination of borehole number and locations for stability analysis of multi-layered slopes using multiple point statistics and information entropy. Can Geotech J 58(11):1669–1689. https://doi.org/10.1139/cgj-2020-0327

Wang X, Wang H, Liang RY (2018) A method for slope stability analysis considering subsurface stratigraphic uncertainty. Landslides 15(5):925–936. https://doi.org/10.1007/s10346-017-0925-5

Wedgwood RJL (1992) AUSTROADS bridge design code. Road Transp Res 1(3):48–54

Wellmann JF, Regenauer-Lieb K (2012) Uncertainties have a meaning: information entropy as a quality measure for 3-D geological models. Tectonophysics 526–529:207–216. https://doi.org/10.1016/j.tecto.2011.05.001

Zhang JZ, Zhang DM, Huang HWf et al. (2022) Hybrid machine learning model with random field and limited CPT data to quantify horizontal scale of fluctuation of soil spatial variability. Acta Geotech 17(4):1129–1145. https://doi.org/10.1007/s11440-021-01360-0

Zhang J, Sun Y, Hu JZ, Huang HW (2023) Assessing site investigation program for design of shield tunnels. Undergr Space 9:31–42. https://doi.org/10.1016/j.undsp.2022.05.002

Zhao T, Wang Y (2019) Determination of efficient sampling locations in geotechnical site characterization using information entropy and Bayesian compressive sampling. Can Geotech J 56(11):1622–1637. https://doi.org/10.1139/cgj-2018-0286

Zhu L, Zhou X, Zhang C (2021) Rapid identification of high-quality marine shale gas reservoirs based on the oversampling method and random forest algorithm. Artif Intell Geosci 2:76–81. https://doi.org/10.1016/J.AIIG.2021.12.001

Žlender B, Jelušič P, Boumezerane D (2012) Planning geotechnical investigation using ANFIS. Geotech Geol Eng 30(4):975–989. https://doi.org/10.1007/s10706-012-9520-7