HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Soils and Foundations 63 (2023) 101269

SOILS AND FOUNDATIONS

www.elsevier.com/locate/sandf

Proposed hybrid approach for three-dimensional subsurface simulation to improve boundary determination and design of optimum site investigation plan for pile foundations

Opeyemi E. Oluwatuyi [a], Rasika Rajapakshage [b], Shaun S. Wulff [c], Kam Ng [d,*]

[a] *Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, 1000 E. University Ave, Laramie, WY 82071-2000, USA*

[b] *Department of Computer Systems Engineering, University of Kelaniya, Sri Lanka*

[c] *Department of Mathematics and Statistics, University of Wyoming, 1000 E. University Ave, Laramie, WY 82071-2000, USA*

[d] *Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, 1000 E. University Ave, Laramie, WY 82071-2000, USA*

Received 9 June 2022; received in revised form 3 November 2022; accepted 11 December 2022
Available online 3 January 2023

Abstract

Geological uncertainty refers to the changeability of a geomaterial category embedded in another. It arises from predicting a geomaterial category at unobserved locations using categorical data from a site investigation (SI). In the design of bridge foundations, geological uncertainty is often not considered because of the difficulties of assessing it using sparse borehole data, validating the quality of predictions, and incorporating such uncertainties into pile foundation design. To overcome these problems, this study utilizes sparse borehole data and proposes a hybrid approach of various spatial Markov Chain (spMC) models and Monte Carlo simulation to predict three-dimensional (3D) geomaterial categories and assess geological uncertainties. The 3D analysis gives realistic and comprehensive information about the site. Characteristics of the proposed hybrid approach include the estimation of transition rates, prediction of 3D geomaterial categories, and simulation of multiple realizations to propagate the uncertainties quantified by information entropy. This proposed hybrid approach leads to specific novelties that include the development of optimal SI plans to reduce geological uncertainty and the determination of geomaterial layer boundaries according to the quantified geological uncertainty. Reducing the geological uncertainties and accurately determining spatial geomaterial boundaries will improve the design reliability and safety of bridge foundations. The hybrid approach is applied to the Lodgepole Creek Bridge project site in Wyoming to demonstrate the application of the hybrid approach and the associated novelties. Outcomes are cross-validated to evaluate the geomaterial prediction accuracy of the hybrid approach.

Keywords: Geostatistics; Geomaterial; Random field; Site investigation; Transition rate matrix

1. Introduction

The performance and reliability of geotechnical structures, like bridge foundations, are largely dependent on supporting geomaterial categories and layers present at a project site (Tran et al., 2018). Borehole exploration is a common site investigation (SI) technique used to obtain

Peer review under responsibility of The Japanese Geotechnical Society.

* Corresponding author.
E-mail addresses: ooluwatu@uwyo.edu (O.E. Oluwatuyi), rasikar@kln.ac.lk (R. Rajapakshage), wulff@uwyo.edu (S.S. Wulff), kng1@uwyo.edu (K. Ng).

https://doi.org/10.1016/j.sandf.2022.101269
0038-0806/

geomaterial and subsurface information. However, it is practically infeasible to collect observations across the whole site with a borehole scheme, which leads to sparse borehole data. This challenge is attributed to the limitation of SI techniques, current SI practice, budget constraints, and unforeseen problems during borehole drilling. Because of these limitations, uncertainty is present at unobserved locations concerning the geomaterial categories, transitions, and combinations thereof. More formally, geological uncertainty refers to the changeability due to one geomaterial layer entrenched in another or the introduction of portions of different geomaterial categories within a homogenous geomaterial mass (Qi et al., 2016). In contrast to the inherent variability of the geomaterial properties, geological uncertainty requires consideration during pile design because the variability of geomaterial properties across different geomaterial categories should be more distinct than the variability of geomaterial properties within the same soil category (Oluwatuyi et al., 2022a, 2022b).

The concept of geological uncertainty is well known and has been studied in the fields of geophysics (Lindsay et al., 2012), hydrology (Elfeki, 2006; Refsgaard et al., 2012), mining (Benndorf and Dimitrakopoulos, 2013; Montiel et al., 2016), and petroleum engineering (Ampomah et al., 2017). Spatial Markov Chains (spMC) have been applied in these applications to simulate geological uncertainty and solve geological problems. For example, the optimization of pump-and-treat systems for the overall pumping rates from the hydraulic barrier is obtained using spMC to minimize geological uncertainty (Pedretti, 2020). Multinomial Categorical Prediction (MCP) consistently produced the highest prediction accuracy from the grab samples data that are interpolated using algorithms from spMC (Prospere et al., 2016). Gao et al. (2016) used a Markov chain transition probability method, like spMC, to simulate spatial distributions converted to categorical distributions in the identification of a representative dataset for long-term monitoring of the Weyburn CO_2-injection enhanced oil recovery site. However, this approach required a large amount of borehole information (900 wells) to define transition rates and probabilities.

In the field of geotechnical engineering, the role of geological uncertainty in layered profiles is initially impossible to study because of the sparse subsurface data available for spatial correlation within the study site (de Marsily et al., 2005). This limitation associated with sparse borehole data creates challenges in geotechnical engineering designs and construction. Although, modeling approaches have been recently developed to accommodate sparse borehole data. These approaches assume that the subsurface domain can be characterized by the Markov property. This property implies that the probability of any future state depends only on the present state and not any past state (Grabski, 2014). For geostatistical applications, the states represent the geomaterial categories and the time steps correspond to the distance in space (Deng et al., 2017). The approaches include the coupled Markov Chain (CMC) model (Deng et al., 2017; Elfeki and Dekking, 2007; Feng et al., 2018), the Markov Random Field (MRF) model (Li and Zhang, 2010; Wang et al., 2019), and the Generalized Coupled Markov Chain (GCMC) model (Deng et al., 2020; Park, 2010; Park et al., 2007). Each approach however possesses limitations. For instance, the CMC model requires borehole data at both sides (far left and right sides) of the 2D study area, yields inaccurate predictions, and underestimates geomaterial in small proportions (Elfeki and Dekking, 2007; Park, 2010). Furthermore, the MRF model requires long computational time, involves complex theory, and is restricted to 2D analysis (Li et al., 2016). In addition, the GCMC model has limited application for large-scale geological simulation due to the difficulty in the direct estimation of the horizontal transition probability matrices (Deng et al., 2020).

Recognizing the limitations of the previous approaches in accommodating sparse data, a hybrid approach that involves the application of various spMC prediction models and Monte Carlo simulation to predict geomaterial categories is proposed in this study. The features of the hybrid approach include the estimation of transition rates, prediction of geomaterial categories at unobserved locations, simulation of multiple realizations to propagate uncertainty, and uncertainty quantification of subsurface stratigraphy using information entropy. To overcome the limitation of analyzing sparse data in spMC, categorical data from boreholes are reported for analysis at every thin depth of 0.075 m of SI. Besides accommodating the sparse borehole data, the proposed hybrid approach is more computationally efficient and allows easy implementation and reproduction in the open-source R program. As a result of these advantages, it is possible to have flexibility in borehole locations within the spatial study site, achieve accurate estimation even in the presence of thin geomaterial layers, and obtain computationally efficient extensions to 3D applications (Feng et al., 2018). The 3D application will give more realistic and comprehensive information about the site.

This study proposes a hybrid approach needed for systematic 3D geomaterial profile modeling and geomaterial layers determination with sparse borehole data through analysis of geological uncertainty. The novelty in the approach involves the development of an optimal SI plan to reduce geological uncertainty and the determination of geomaterial layer boundaries according to geological uncertainty. The development of the optimum SI plan will give engineers the ability to quantitatively communicate and convince their clients of the need for an additional borehole to improve the design and construction of foundation systems. The determination of geomaterial boundaries using quantified geological uncertainty is an important advancement. The true boundaries of a site will be impossible to determine without drilling boreholes at every location. The nonlinear spatial boundaries obtained from the proposed hybrid approach will provide a better alternative to the assumed two-dimensional (2D) linear

interpolation method that may under or overestimate the thickness of the different geomaterial categories. These spatial boundaries also provide a means to combine geological uncertainties with other forms of uncertainties to predict geomaterial properties for the reliability-based limit state design (Huffman et al., 2015). The methodology and analyses presented in this study can be reproduced by engineers for the design of bridge foundations and other geotechnical structures at different site locations. The hybrid approach is demonstrated using the Lodgepole Creek bridge project site in Wyoming, and results are cross-validated to evaluate model prediction accuracy.

The paper is organized as follows. First, the methodology underlying the proposed hybrid approach is presented to generate estimations, predictions, simulations, and uncertainty measures. Next, a case study is described to illustrate the 3D analysis with the hybrid approach. Then, the results from the 3D analysis for the bridge project site are presented in subsequent discussions. Finally, conclusions and recommendations are provided.

2. Methodology

2.1. Background

The methodology in this section describes the proposed hybrid approach which uses various spMC models and Monte Carlo simulation to estimate, predict, and simulate the categorical geomaterials at unobserved locations of the site. The models are fit using the spMC package (Sartore, 2019, 2013; Sartore et al., 2016) in the R program (R Core Team, 2016). Monte Carlo simulation is also integrated into the package. This package "provides several functions to deal with categorical spatial data and continuous lag Markov Chains" (Sartore et al., 2016). This package has the following advantages (i) it is implemented in the open-source program of R so that it is widely available to practitioners, (ii) it has high-performance computational (HPC) techniques to aid in computational efficiency, and (iii) it uses recent advances for conducting simulations. The required inputs for analysis include a vector of categorical data (primary data) obtained from the boreholes and the respective 3D location coordinates of the boreholes (Sartore, 2019). A brief overview of the terminology, equations, and procedures for the proposed hybrid approach are provided here to outline the necessary background for the paper. To resolve the issue of sparse data required to define transition rates, the data from the boreholes are reported at a depth of 0.075 m. The transition rate matrices are computed as described in the next section.

2.2. Estimation of transition rate matrices

Consider a spatial location $\boldsymbol{s}$ and let $Z(\boldsymbol{s})$ denote the categorical random variable for the state or geomaterial at that location. In geostatistics, measurements of Z are available at observed locations and predictions of Z are required at unobserved locations. Thus, $Z(\boldsymbol{s})$ forms a random field across all locations $\boldsymbol{s}$ within a study site (Bivand et al., 2013). Now, consider two locations $\boldsymbol{s}$ and $\boldsymbol{s}+\boldsymbol{h}$where $\boldsymbol{h}$ is the multidimensional lag and $\|\boldsymbol{h}\|$ denotes 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 categorical state z_i to categorical state z_j is

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

where $i, j = 1, 2, \cdots, n_S$, and n_S is the number of states. The collection of transition probabilities $t_{ij}(\boldsymbol{h})$ in Equation (1) across all $i, j$is given by the $n_S \times n_S$ transition probability matrix, $\boldsymbol{T}(\boldsymbol{h})$ as

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

where $\boldsymbol{R}_{\boldsymbol{h}}$ is the transition rate matrix which depends on the direction given by the lag $\boldsymbol{h}$(Sartore et al., 2016). The transition rate illustrates the immediate rate at which the Markov chain transitions between geomaterial categories. These transition rates help understand the subsurface stratigraphy and in obtaining the corresponding predictions for a site.

Carle and Fogg (1997) propose an approximation to express the transition rate matrix $\boldsymbol{R}_{\boldsymbol{h}}$ ($\boldsymbol{R}_{\boldsymbol{e}_k}$)in Equation (2) as

$$\boldsymbol{R}_{\boldsymbol{e}_k} = \operatorname{diag}\Big(\{\bar{L}_{i,\boldsymbol{e}_k}\}\Big)^{-1}(\boldsymbol{F}_{\boldsymbol{e}_k} - \boldsymbol{I}) \tag{3}$$

where $k = 1, 2, \cdots, d$ for d dimensions, $\boldsymbol{e}_k$is the standard basis vector for the direction indexed by k, $\bar{L}_{i,\boldsymbol{e}_k}$ is the mean stratum thicknesses of state i along the direction $\boldsymbol{e}_k$, diag denotes a diagonal matrix with entries $\bar{L}_{i,\boldsymbol{e}_k}$, and $\boldsymbol{F}_{\boldsymbol{e}_k}$ is the transition probability matrix consisting of probabilities for consecutive elements with the same geomaterial category along the direction $\boldsymbol{e}_k$ (Sartore et al., 2016). The transition rate ($\boldsymbol{r}_{ij,\boldsymbol{h}}$) in row i and column j of the transition rate matrix $\boldsymbol{R}_{\boldsymbol{h}}$ in Equation (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 Equation (3), $\boldsymbol{r}_{ij,\boldsymbol{h}}$ is non-positive when $i = j$, and $\boldsymbol{r}_{ij,\boldsymbol{h}}$ is non-negative when $i \neq j$ (Sartore et al., 2016). The components $\boldsymbol{r}_{ij,\boldsymbol{e}_k}$ in Equation (4) can be estimated using the Maximum Entropy (ME) method, which estimates the components of $\boldsymbol{F}_{\boldsymbol{e}_k}$ by plugging in $f^*_{ij,\boldsymbol{e}_k}$into Equation (3) upon convergence of the following iterative algorithm.

1. Initialize $f_{i,\boldsymbol{e}_k} = \frac{p_i}{\bar{L}_{i,\boldsymbol{e}_k}}$ for $i = 1, 2, .., n_S$.
2. Compute $f^*_{ij,\boldsymbol{e}_k} = f_{i,\boldsymbol{e}_k} f_{j,\boldsymbol{e}_k}$ for $, j = 1, 2, .., n_S$.
3. Compute $f_{i,\boldsymbol{e}_k} = \frac{p_i \sum_{s=1}^{n_S}\sum_{j\neq s}^{n_S} f^*_{sj,\boldsymbol{e}_k}}{\bar{L}_{i,\boldsymbol{e}_k}\sum_{j\neq i}^{n_S} f^*_{ij,\boldsymbol{e}_k}}$.
4. Repeat steps 2 and 3 until convergence.

where p_i is the proportion in state i (Sartore, 2013; Sartore et al., 2016).

2.3. Subsurface prediction

Geological modeling is conducted to predict subsurface characteristics or geomaterial categories at an unobserved location $(\boldsymbol{s}_0)$ using the categorical data from a site $(\boldsymbol{s}_l, l = 1, \cdots, n)$.

For prediction, the conditional probability of interest involves the unknown location and the combination of the data rather than just the two locations as in Equation (1). The conditional probability of interest is then given by

$$q_j(\boldsymbol{s}_0) = \Pr\left(Z(\boldsymbol{s}_0) = z_j \middle| \bigcap_{l=1}^{n} Z(\boldsymbol{s}_l) = z(\boldsymbol{s}_l)\right) \tag{5}$$

where $z(\boldsymbol{s}_l)$ is the value of the random variable $Z(\boldsymbol{s}_l)$at the observed location $\boldsymbol{s}_l$ (Sartore et al., 2016).

To predict geomaterial category j at an unobserved location $(\boldsymbol{s}_0)$, an approximation of the conditional probability in Equation (5) is needed. Various algorithms can be considered for the approximation, namely Kriging, Path Algorithms, and Multinomial Categorical Prediction (Sartore et al., 2016). The Fixed and Random Path Algorithms (Li, 2007; Li and Zhang, 2007) are not pursued in this study due to concerns that these methods are not correctly implemented in the spMC package (Li and Zhang, 2015).

The Kriging methods are subdivided into Indicator Kriging (IK) and Indicator CoKriging (CK). Data in geotechnical engineering are sorted based on their acquisition and usage into borehole data, geological mapping data, and property data (Zhang and Zhu, 2018). In this study, Kriging is targeted solely at the borehole data as the primary data which includes the vector of geomaterial categories present in the borehole and its 3D location coordinates. The two Kriging methods IK and CK employed in this study are for comparison purposes with respect to this primary data. For both Kriging methods, the conditional probability in Equation (5) is approximated using the weighted linear combination given by

$$q_j(\boldsymbol{s}_0) \approx \sum_{l=1}^{n} \sum_{i=1}^{n_S} w_{ij,l} c_{il} \tag{6}$$

where c_{il} is an indicator function based on the transition probability estimated in Section 2.2 taking the value 1 when $z(\boldsymbol{s}_l) = z_i$ and 0 otherwise, and $w_{ij,l}$ are the weights (Zhang and Zhu, 2018). Using the $n_S \times n_S$ transition probability matrix in Equation (2), the weights can be found using a system of equations given by

$$\begin{bmatrix} \boldsymbol{T}(\boldsymbol{s}_l - \boldsymbol{s}_l) & \cdots & \boldsymbol{T}(\boldsymbol{s}_n - \boldsymbol{s}_l) \\ \vdots & \vdots & \vdots \\ \boldsymbol{T}(\boldsymbol{s}_l - \boldsymbol{s}_n) & \cdots & \boldsymbol{T}(\boldsymbol{s}_n - \boldsymbol{s}_n) \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_l \\ \vdots \\ \boldsymbol{W}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}(\boldsymbol{s}_0 - \boldsymbol{s}_l) \\ \vdots \\ \boldsymbol{T}(\boldsymbol{s}_0 - \boldsymbol{s}_n) \end{bmatrix} \tag{7}$$

where the $n_S \times n_S$matrix $\boldsymbol{W}_l$ has (i, j) entry $w_{ij,l}$ for observation with index l and $\boldsymbol{T}(\boldsymbol{s}_l - \boldsymbol{s}_n)$represents the transition rate matrix in Equation (2) from the location $\boldsymbol{s}_l$ to location $\boldsymbol{s}_n$. Unbiased constraints are placed on the weights in which IK uses the weights in Equation (8) and CK uses the weights in Equation (9)

$$\sum_{l=1}^{n} w_{ij,l} = 1 \tag{8}$$

$$\sum_{l=1}^{n} \boldsymbol{W}_l = I \tag{9}$$

where I in Equation (9) is the $n_S \times n_S$identity matrix (Carle and Fogg, 1996; Myers, 1982). While in theory, CK should be the best algorithm for estimating conditional probabilities, CK may not provide a more accurate estimation than other procedures for categorical-based data (Allard et al., 2011; Goovaerts, 1994).

The Multinomial Categorical Prediction (MCP) method initially proposed by Bogaert (2002), and improved by Allard et al. (2011), is based on the Bayesian maximum entropy where the conditional probability in Equation (5) is approximated using

$$q_j(\boldsymbol{s}_0) \approx \frac{p_i \prod_{l=1}^{n} t_{ik_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)}{\sum_{i=1}^{n_S} p_i \prod_{l=1}^{n} t_{ik_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)} \tag{10}$$

where p_i is the proportion in category i and $t_{ik_l}(\boldsymbol{s}_0 - \boldsymbol{s}_l)$represents the transition probability as defined in Equation (1) from the state z_i to state z_k for observation with index l (Sartore et al., 2016). Allard et al. (2011) indicate that the MCP method can reproduce complex and asymmetrical patterns.

Predictions can be obtained based on the approximations of the conditional probabilities for all geomaterial categories in the gridded elements of the random field. The predicted geomaterial category for the spatial location corresponds to that category with the highest conditional probability. It is possible to use all the data or only n_N of the nearest neighbors in the approximations in Equations (6) and (10) to aid in the computational efficiency. Computational efficiency is especially important when it comes to the simulation process. Thus, this study uses the recommended default of 12 nearest neighbors (Sartore et al., 2016).

2.4. Subsurface simulation

Monte Carlo simulation is performed by repeating the CK, IK, and MCP methods n_M times to approximate the conditional probabilities, $q_j(\boldsymbol{s}_0)$ in Equation (5) (Wang and Cao 2013). The number of simulations (n_M) is determined through a sensitivity analysis. These simulations are needed to propagate the uncertainty in the predictions as the predicted geomaterial categories can vary at the unobserved spatial locations. The multiple simulated conditional probabilities are evaluated and used in the quan-

tification of uncertainty. The proposed hybrid approach is probabilistic such that uncertainties in the conditional probability values are quantified using information entropy which is described in the next section.

2.5. Information entropy

To quantify the geological uncertainty on the subsurface predictions, randomized realizations are generated through simulation. The simulation is conducted by random selection of a category or state according to the approximated probability in Equation (5). The uncertainty calculation is then based on information entropy across the simulations.

Information entropy is a quantitative way of describing the associated geological uncertainty in a subsurface prediction. The general form for information entropy is given by

$$H = -\sum_{j=1}^{n_s} p_j \log_{n_s} p_j \tag{11}$$

where p_j is the conditional probability of category j for $j = 1, 2, \cdots, n_s$ (Wellmann and Regenauer-Lieb, 2012). Recall that in section 2.2 n_S has earlier been described as the number of geomaterial states or categories. The minimum of H is 0 when $p_h = 1$ and $p_j = 0$ for all $j \neq h$ using the convention $\log_{n_s} 1 = 0$. The maximum of H is $\log_{n_s} n_s = 1$. Thus, the information entropy is low (low uncertainty) when one p_j is near 1 and is high (high uncertainty) when the p_j are close to equal in value across the geomaterial categories.

The information entropy in Equation (11) needs to be modified for this application. That is, H is calculated at a gridded spatial location $\boldsymbol{s}_0$ using an approximation of the conditional probability $q_{j,m}(\boldsymbol{s}_0)$ in Equation (5) for one simulation indexed by m out of the multiple simulations n_M. The corresponding equation is given by

$$H_m(\boldsymbol{s}_0) = -\sum_{j=1}^{n_s} q_{j,m}(\boldsymbol{s}_0) \log_{n_s} q_{j,m}(\boldsymbol{s}_0) \tag{12}$$

The location mean information entropy, $\bar{H}(\boldsymbol{s}_0)$, is obtained at a gridded spatial location by averaging the information entropy values in Equation (12) across the n_M simulations. Overall mean information entropy is obtained across the site, by averaging $\bar{H}(\boldsymbol{s}_0)$ across all the spatial locations in the site to give $\bar{\bar{H}}$.

2.6. Validation

Validation is the process by which the accuracy of the predictions obtained from a model, such as from a given spMC model, is evaluated on a new set of data (Kutner et al., 2004). True 3D ground information describing the geomaterial profile throughout the study site is not available in actual geotechnical investigations. Cross-validation is a viable alternative that can be used to assess prediction performance when true ground information is unavailable (Kutner et al, 2004). Cross-validation is effective for this purpose because (1) model fitting is conducted on training data without the use of the testing data, (2) model evaluation is based on the testing data only, thereby treating such data as unseen by the model fitting process, (3) the process is repeated throughout all testing datasets (folds) which allows for ground truthing evaluation independent of the model fitting process. The three methods, namely IK, CK, and MCP, for predicting the geomaterial category from the spMC model are compared in this study. The objective of cross-validation for this study is to determine which of the methods is most effective and accurate in geomaterial prediction. Validation of the model can provide insights into the practical viability of the resulting predictions. However, the assessment of geological uncertainty is site-specific, so using data from another site would be inappropriate. Thus, the validation process in this study involves k-fold cross-validation, which is a commonly used technique when validation data is neither practical nor available (Kutner et al., 2004).

With k-fold cross-validation, CV(k), the data in one part of the fold forms the validation data, the remaining data forms the training data, and this process is repeated across the k folds (James et al., 2015). Usually, the folds are randomly selected. However, in this application, it makes more sense to use the data in each borehole as a fold. Thus, the data for a particular borehole, known as the validation data, is removed from the overall dataset, known as the training data, in order to generate predictions from the spMC model. The observed geomaterial categories for the validation data are compared to the predicted geomaterial categories obtained from the predictions based on the training data. The percentage match is computed for each validation dataset or borehole to obtain a cross-validation index. This percentage match is calculated for each simulation and then averaged across the simulations to obtain the mean percentage match.

The proposed hybrid approach described in this section is applied to an actual case study as described in the next section. It is important to realize that a major advantage of the proposed hybrid approach is its generalizability to any site, which has available information on the categorical states consisting of the geomaterials along with the respective 3D location coordinates. With this information, it is possible to carry out a complete and thorough analysis of geological uncertainty. A workflow of the methodology used in this paper is shown in Fig. 1 to aid in the implementation of the proposed hybrid approach. The key steps involve assigning data to the discretized (gridded) elements, simulating the predictive surface with the estimated transition rates, multiple simulations using Monte Carlo simulation, quantifying the geological uncertainty using the information entropy, and performing the above validation procedure.

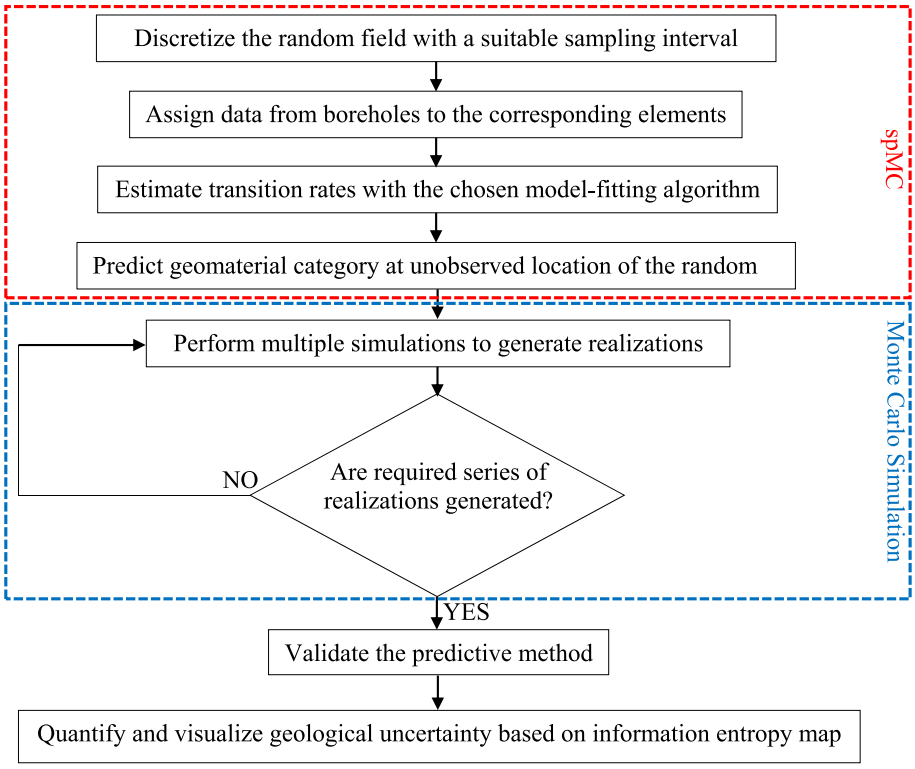

Fig. 1. Flowchart showing the geological profile modeling and uncertainty quantification using the proposed hybrid approach.

3. Case study

The case study consists of an actual project to replace an existing five-span bridge founded on a reinforced concrete pile over Lodgepole Creek. The area of construction is situated within Pine Bluffs to Albin, North Section, WY 215, Laramie County, Wyoming, USA. A foundation investigation report by the Wyoming Department of Transportation (WYDOT) consists of six boreholes shown in Fig. 2. The arrows pointing to each borehole (BH) in Fig. 2 show the position and ground elevation of the boreholes. The ground surface elevation, groundwater-surface, and geological profile as estimated by the engineering geologist are shown in Fig. 3. Altogether six boreholes are plotted in Fig. 4 to form the data to be analyzed for this study. Four geomaterial strata are identified on the project site through the drilled boreholes. The overburden soils consist of a mixture of sand with silt and gravel, sand, and silt, all underlain by an intermediate geomaterial (IGM) layer of sandy siltstone. An American Association of State Highway and Transport Officials (AASHTO) and Unified Soil Classification System (USCS) classification of the identified geomaterial strata is presented in Table 1. The random field covers the project site of + 60 m in the longitudinal distance, +30 m in the transverse distance, and −30 m deep. Negative values are used for describing depth below the ground surface.

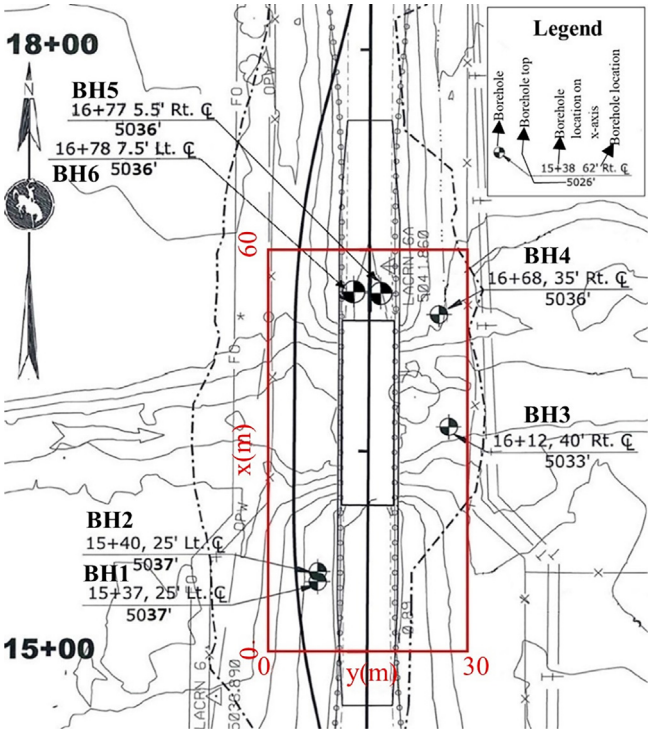

Fig. 2. Aerial view of borehole layout (red rectangular lines indicate the study area) (Ng et al., 2023).

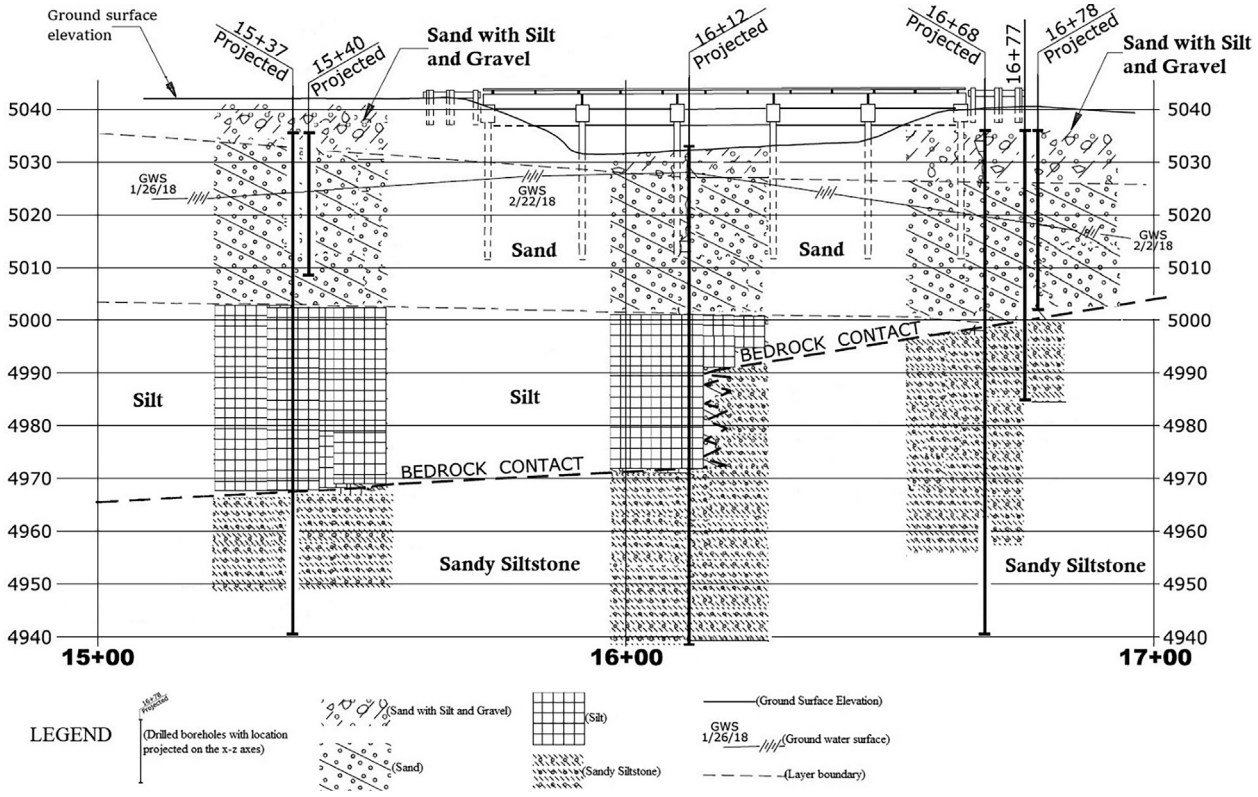

Fig. 3. Projected x-z section view of the borehole layout (Ng et al., 2023).

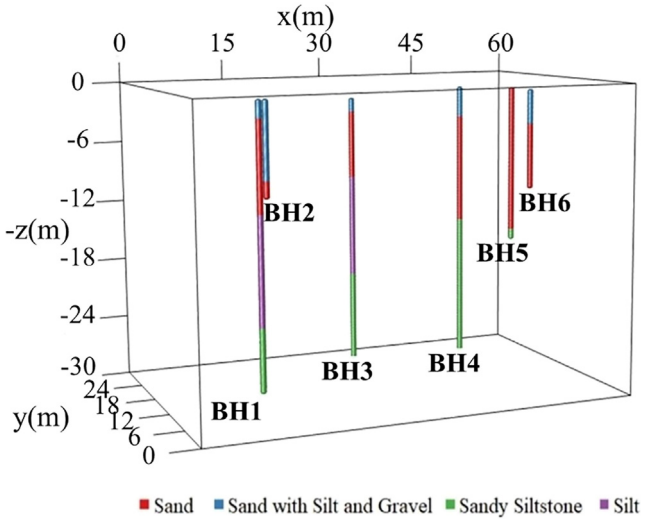

Fig. 4. 3D view of the borehole layout (Ng et al., 2023).

Table 1
Geomaterial classification.

Geomaterial label	AASHTO classification	USCS* classification
Sand with silt and gravel	A-1-b to A-2–4(0)	GM, SP-SM
Sand	A-1-b to A-2–7(0)	SW, SM
Silt	A-4(0)	ML
Sandy siltstone*	A-5(11)	ML

AASHTO-American Association of State Highway and Transportation Officials; USCS-Unified Soil Classification System; and Soil classification is on cuttings from the geomaterial.

4. Results

4.1. Estimation of transition rate matrices

The estimated transition rate matrices in 3D for the borehole data are presented in Table 2 to provide an understanding of the subsurface stratigraphy of the site. The transition rate of a geomaterial category to itself is a negative value so that the rows of the transition rate matrix sum to zero. For example, the estimated transition rate of sandy siltstone to itself in the transverse direction is −0.0703. The passage intensity (transition rate magnitude of a geomaterial to itself) is always higher than the transition intensity (transition rate magnitude of a geomaterial to another). For example, the estimated transition rate of sandy siltstone to itself in the transverse direction (0.0703) is higher in magnitude than that of sandy siltstone to silt (0.0403). The estimated transition rates with positive values in the longitudinal (x-axis) distance are smaller in magnitude because of their longer length (+60 m) when compared to the rates in the depth (z-axis) and transverse (y-axis) direction (30 m). The proportions of the geomaterial categories as reflected by the borehole data are also estimated alongside the transition rate matrices. These proportions are 0.1626, 0.3952, 0.1944, and 0.2478 for sand with silt and gravel, sand, silt, and sandy siltstone, respectively. From the estimation transition rate, a conditional transitional probability between 0 and 1 is then generated using the IK, CK, and MCP methods for each geomaterial category

Table 2
3D estimated transition rate matrices obtained using the maximum entropy method (Ng et al., 2023).

X-direction estimated transition rate matrix				
Geomaterial category	Sand with silt and gravel	Sand	Silt	Sandy siltstone
Sand with silt and gravel	−0.0087	0.0040	0.0022	0.0025
Sand	0.0032	−0.0106	0.0034	0.0040
Silt	0.0020	0.0039	−0.0084	0.0025
Sandy siltstone	0.0021	0.0041	0.0023	−0.0085
Y-direction estimated transition rate matrix				
Geomaterial category	Sand with silt and gravel	Sand	Silt	Sandy siltstone
Sand with silt and gravel	−0.0876	0.0194	0.0491	0.0191
Sand	0.0091	−0.0455	0.0262	0.0102
Silt	0.1179	0.1342	−0.3840	0.1319
Sandy siltstone	0.0140	0.0160	0.0403	−0.0703
Z-direction estimated transition rate matrix				
Geomaterial category	Sand with silt and gravel	Sand	Silt	Sandy siltstone
Sand with silt and gravel	−0.0781	0.0320	0.0211	0.0250
Sand	0.0148	−0.0386	0.0109	0.0129
Silt	0.0131	0.0147	−0.0392	0.0114
Sandy siltstone	0.0143	0.0161	0.0106	−0.0410

at each gridded location. To illustrate the geomaterial category prediction, the probabilistic parameter (i.e., the highest conditional probability at every gridded location of the random field) as shown in Fig. 5 is selected to obtain the most probable geomaterial category prediction. Imagine a gridded location on the random field after the analysis, the conditional probabilities of having sand with silt and gravel, sand, silt, and sandy siltstone at a such gridded location are estimated at 0.10, 0.25, 0.55, and 0.10 respectively. The 0.55 probability value will be plotted for this gridded location in Fig. 5. The algorithm will thereafter choose 'silt' as the predicted geomaterial category for that gridded location because it has the highest probability of occurring at that location.

4.2. Subsurface stratigraphy predictions and uncertainty quantification

One hundred simulations of the most probable geomaterial category prediction for the subsurface are generated using the IK, CK, and MCP methods. Each simulation produces predicted geomaterial categories across the 3D study region based on the available borehole data. The 100 simulations are averaged and its geomaterial predictions for IK, CK, and MCP are shown in Fig. 6. The figure shows that for the overburdened soil, which is at the uppermost part of the subsurface, sand with silt & gravel is mostly predicted followed closely by sand, and silt. At the lower depth, which is the bedrock, the predicted geomaterial is sandy siltstone. The averaged value from the simulations illustrates the similarity between the IK and CK methods. It also illustrates the asymmetry captured by the MCP method, particularly for the Silt layer. The estimated mean thickness of each geomaterial layer in the MCP predicted subsurface is 3.84 m for sand with silt and gravel, 7.77 m for sand, 7.65 m for silt, and 7.31 m for sandy siltstone.

Several simulations are needed to propagate the uncertainty in the predictions as the predicted geomaterial categories can vary at the unobserved spatial locations. A sensitivity analysis is first conducted to study the effect of the number of simulations on the standard deviation of the overall mean information entropy. Fig. 7 shows a plot of standard deviation with 10, 50, 100, 200, and 500 simulations using MCP. This figure shows a dramatic reduction in the standard deviation at 100 simulations with only a small reduction (less than 1 %) after 100 simulations. Thus, 100 simulations appear to be adequate in this study as additional simulations do not appear to justify the additional computational time. Models, like CMC and MRF (earlier described in the introduction section), require more simulation samples (more than 10,000) to converge. The proposed hybrid approach, consisting of spMC and MCS, does not require as many simulations.

An information entropy map can be produced by plotting the location mean information entropy at the location $\boldsymbol{s}_0$ as given by $\bar{H}(\boldsymbol{s}_0)$. Fig. 8 displays this map for the CK, IK, and MCP methods. The color bar labels on the right-hand side of the figures represent the uncertainty quantified at each element. It should be noted that there is more uncertainty associated with CK and IK than with MCP as is evident from the figures with the intensity of the yellow to red colors in CK and IK compared to MCP. The overall mean information entropies are 0.196 for CK, 0.200 for IK, and 0.014 for MCP.

4.3. Validation of simulated subsurface predictions

As discussed in Section 2.6, k-fold cross-validation is used to evaluate the prediction accuracy of the simulation approaches for the spMC model. In this study, only six boreholes are used, because data from boreholes for geotechnical site investigation is costly and difficult to obtain. Nevertheless, the proposed hybrid approach is well

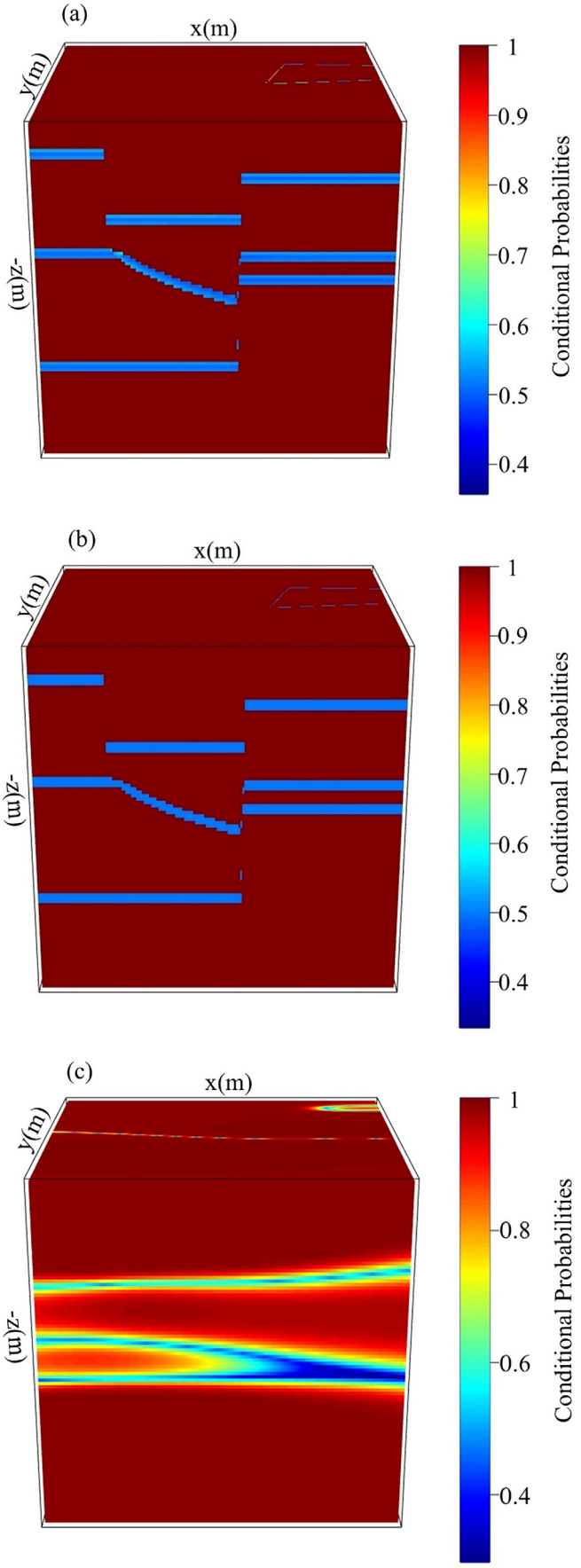

Fig. 5. Highest conditional probability at grid positions on the random field as estimated by the (a) IK, (b) CK, and (c) MCP methods (Ng et al., 2023).

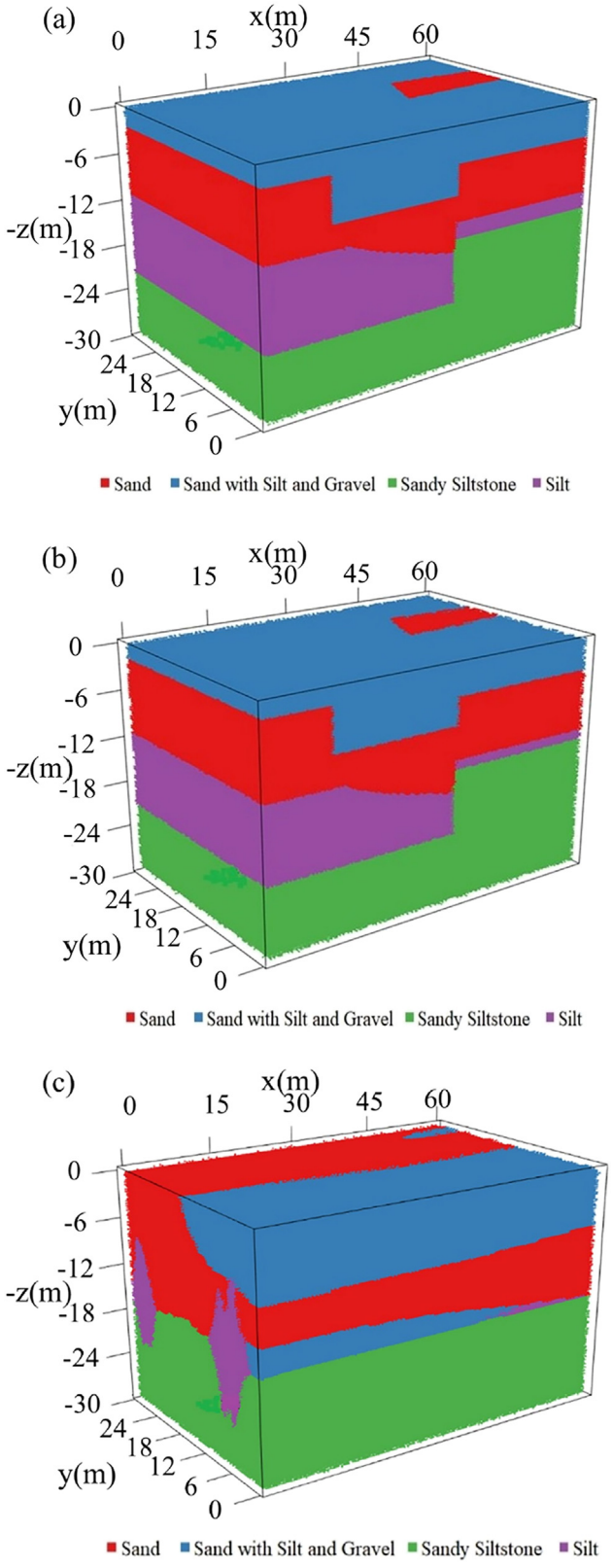

Fig. 6. Geomaterial categories as predicted by (a) IK, (b) CK, and (c) MCP (Ng et al., 2023).

suited for modeling such sparse borehole data. The given borehole data provide enough information to estimate the transition matrix in the horizontal direction. The relia-

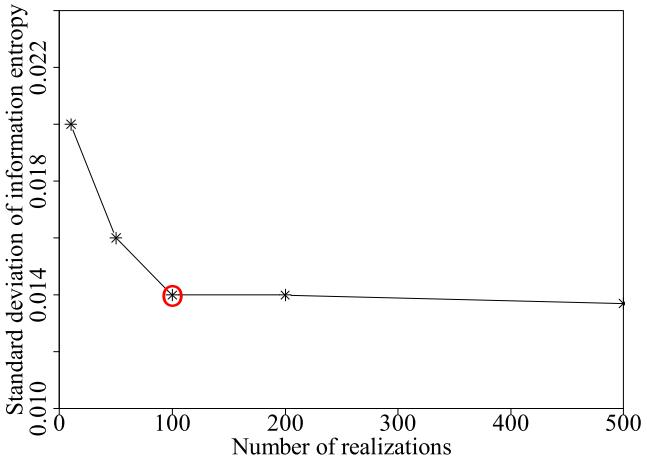

Fig. 7. Sensitivity analysis on the number of simulations with MCP based upon the standard deviation of the overall mean information entropy (Ng et al., 2023).

bility of the comparison among the three methods is evaluated through the process of cross-validation and the results are compared with another related study. The mean percentage match for the boreholes across the simulations is evaluated for the different methods IK, CK, and MCP. The results are shown in Table 3.

The mean percentage match for BHs 1 and 2 is poor with less than 41 % match. Notice that these two boreholes are quite close together as shown in Figs. 2 – 4. Despite their proximity, there is quite a discrepancy between the observed thicknesses of the sand layer and the sand with silt and gravel layer. In addition, BH2 has a small depth and lacks observed information on the remaining two layers (silt and sandy siltstone). The rest of the boreholes have a high mean percentage match, with marginal percentage matching for BH5. The cross-validation index value indicates that MCP appears to be the best predictive simulation algorithm with a mean percentage match of 65 % across all boreholes. The higher prediction accuracy in MCP over CK and IK is attributed to less uncertainty associated with the MCP method. In a previous study by Prospere et al. (2016), interpolation methods from spMC are used in the substrate mapping of three rivers in Jamaica namely Gayle, Black, and Salt Springs Rivers. For all three rivers, accuracy assessments on the prediction methods revealed the most accurate predictions are made consistently by the MCP method, which is in line with the present study. Mean prediction accuracies obtained from using the MCP method are 70.3 ± 9.2 %, 65.8 ± 10.4 %, and 63.0 ± 10.3 % for Gayle, Black, and Salt Springs Rivers, respectively (Prospere et al., 2016). For this study, an MCP geomaterial category prediction at an unobserved location within the study area has a 65 % mean prediction accuracy. This degree of accuracy is especially encouraging considering the sparse amount of borehole data, which is even more severe under the cross-validation scheme in which a particular borehole is removed. Oluwatuyi et al. (2022b) also showed that the MCP method tended to perform better than the IK and CK methods at predicting the observed

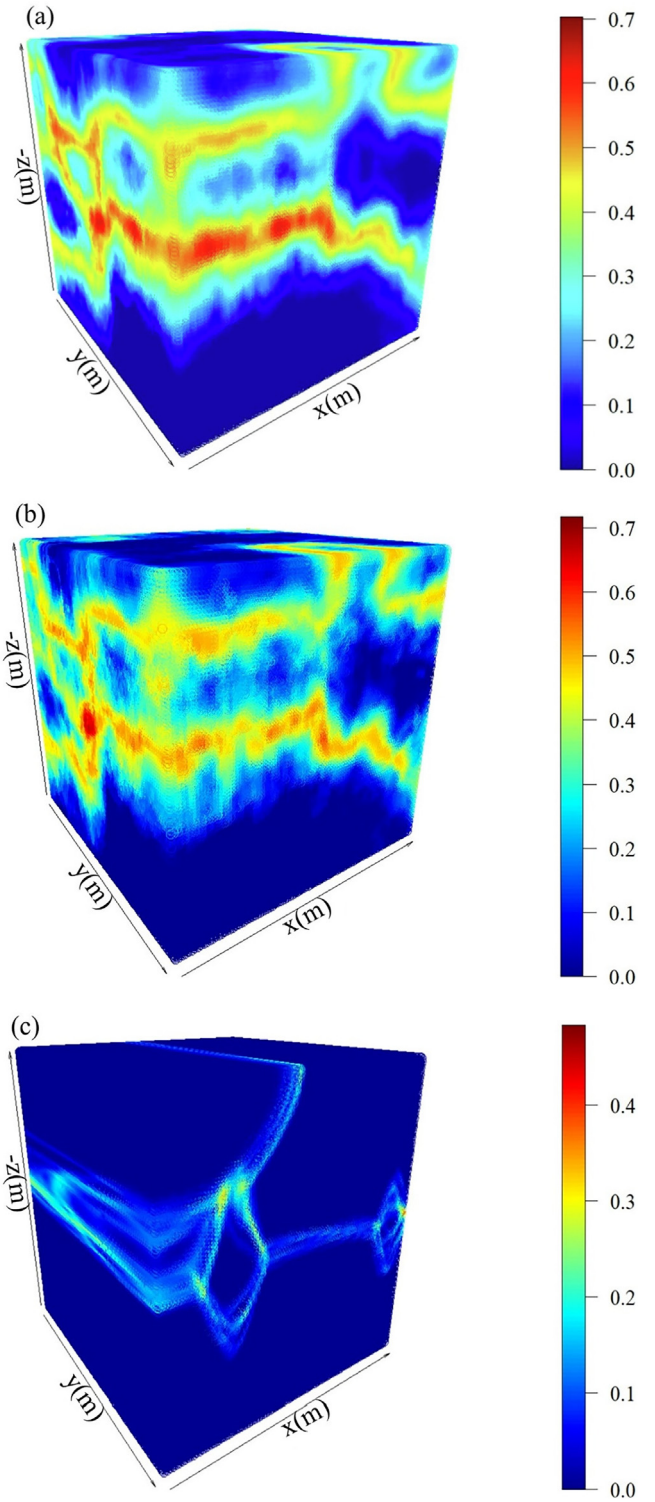

Fig. 8. Map of the location mean information entropy using (a) IK, (b) CK, and (c) MCP (Ng et al., 2023).

Table 3
Summary of the cross-validation process (Ng et al., 2023).

Borehole number and location (x and y coordinates in parenthesis)		Percentage match						
		BH1 (37,25)	BH2 (40,25)	BH3 (112,90)	BH4 (168,85)	BH5 (177,55.5)	BH6 (178,42.5)	Mean Across Boreholes
spMC Model method	IK	19.56	40.62	61.29	67.82	49.12	70.59	51.50
	CK	20.65	37.50	61.29	70.11	49.12	70.59	51.54
	MCP	30.43	40.62	88.06	96.55	57.37	77.64	65.11
Borehole length (m)		27.60	8.40	27.90	28.50	15.60	10.20	NA

BH1-Borehole with number and location; NA-Not applicable.

borehole information. Thus, the MCP algorithm is recommended and used exclusively in the rest of the study.

The computational efficiency of the different simulation algorithms in spMC has been studied and evaluated by Sartore et al., (2016). While the run times of the hybrid approach used are not recorded in this study due to numerous subsequent calculations, this study can confirm that based on its study site and primary data, MCP is more computationally efficient than IK and CK.

4.4. Site investigation plans to reduce geological uncertainty

For this study site, the foundation locations are the north abutment (N. ABUT.), the south abutment (S. ABUT.), and the center pier (C. PIER.). The sum of the location mean entropy, $\bar{H}(\boldsymbol{s}_0)$, is taken across locations $\boldsymbol{s}_0$along the geomaterial depth (z) of a similar geomaterial layer for a given longitudinal and transverse (x,y) direction (plan view). The sum of the location mean entropies for each layer are mapped in Fig. 9 showing the areas around the bridge pile foundation. It should be noted that the mean information entropy values estimated in each of the four geomaterial layers, namely sand with silt and gravel, sand, silt, and sandy siltstone are 0.006, 0.017, 0.025, and 0.008, respectively. The mean information entropy for silt is the highest, which may be due to the thinness of the layer and the low number of gridded elements. The mean of these four values equals 0.014, which is the same as the overall mean information entropy of the whole site for MCP as shown in Fig. 8(c).

The design of various borehole layout schemes is done in the SI phase. Recall, that the original layout scheme is based upon six boreholes with (x,y) coordinates given in Table 3. Variations of the original layout scheme are considered in this study to demonstrate the importance of borehole location and to illustrate how to obtain an optimal SI plan. Table 4 shows these variations where the borehole scheme identification represents the number of boreholes. The original borehole scheme corresponds to scheme 6 in Table 4. Boreholes marked with "X" are the exact boreholes from the original layout with their coordinates listed in the column heading. Boreholes marked with "RBH" are recommended boreholes as alternatives to the original layout scheme. The purpose of the recommended RBH boreholes is to demonstrate the importance of strategic borehole location in reducing geological uncertainties as measured by entropy. The boreholes associated with RBH are recommended at these locations, which are different from the original locations, to reduce the geological uncertainty associated with a site investigation plan. For comparable analysis of the different borehole schemes, it is assumed all boreholes are drilled to the full 30 m depth (z) in the study site.

It is possible to evaluate these schemes in terms of geological uncertainty, as measured by information entropy, for all geomaterials or a particular geomaterial of interest. In this study, evaluation and reduction of uncertainty are specified to the sandy siltstone because most H-piles used in bridge construction at the study site are end-bearing piles. Reduction of whichever measure of geological uncertainty can be accomplished with the addition of boreholes at the subsurface (x,y) locations with the highest values of information entropy. For example, consider borehole scheme 6v as an alternative to borehole scheme 6. Two recommended boreholes 'RBH2′ and 'RBH6′ are added as alternative replacements to BH2 and BH6 in the original design. The locations of RBH2 and RBH6 are determined by the (x,y) coordinates that have the highest total location mean entropy in sandy siltstone. These coordinates are marked with a red square in the upper right (RBH2) and lower right (RBH6) of Fig. 9 (d). The overall mean information entropies of the various borehole schemes are displayed in Fig. 10. Thus, the geological uncertainty for the original borehole scheme 6 is 0.011 and drops to 0.006 using borehole scheme 6v. Generally, the geological uncertainty of a site can be reduced with the addition of more boreholes. However, borehole location also is important as demonstrated in the comparison between schemes 6 and 6v. From Fig. 10, the overall mean information entropy using borehole scheme 4 is 0.007, which increases to 0.01 with the additional borehole in scheme 5a. With the repositioning of the additional borehole in scheme 5b, the overall mean information entropy is at 0.007, which is the same as that of scheme 4. In fact, scheme 4 has lower overall mean information entropy than both scheme 5a and scheme 6. While borehole scheme 6v has the smallest overall mean information entropy, the SI plan offered a minimal improvement over borehole scheme 4 and scheme 5b.

In convincing the client of the need for an additional recommended borehole, Figs. 9 and 10 with Table 4 could

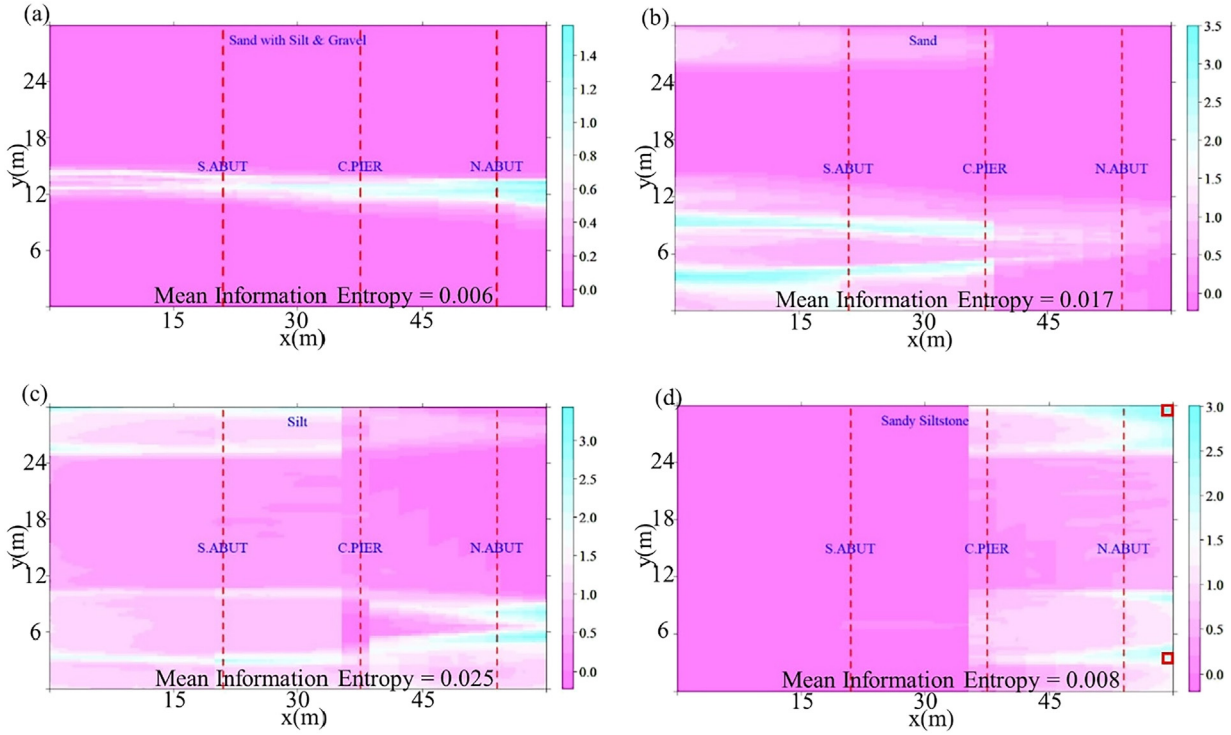

Fig. 9. Cross-section (x,y) for each geomaterial layer: (a) sand with silt & gravel, (b) sand, (c) silt, and (d) sandy siltstone using MCP to display the sum of the location mean entropy across the depths (z) (Ng et al., 2023).

Table 4
Borehole schemes to assess uncertainty reduction (Ng et al., 2023).

Borehole scheme	Borehole number and location (x,y) coordinates in parenthesis					
	BH1 (11.1,7.5)	BH2 (12,7.5)	BH3 (33.6,27)	BH4 (50.4,25.5)	BH5 (53.1,16.65)	BH6 (53.4,12.75)
4	X		X	X	X	
5a	X		X	X	X	X
5b	X	X	X	X	X	
6	X	X	X	X	X	X
6v	X	RBH2 (60,30)	X	X	X	RBH6 (60,3)

BH - Borehole with number and location; RBH - Recommended borehole with number and location; X - Borehole present.

be used as possible communication tools. The possible way to know the geomaterials at these locations with high uncertainties and in turn reduce the geological uncertainty is to recommend boreholes at these locations.

4.5. Boundary description of geomaterial layers

The characterization of geomaterial layer boundaries is critical to the design of an adequate deep foundation system. In this section, the spatial boundaries between the different geomaterial layers are estimated. The conventional practice, apart from the local geological experience of the project site, is to determine the boundaries through linear interpolation of observed geomaterials from well-spaced boreholes in 2D. This linear interpolation method considers the site-specific geological history as estimated by an experienced engineer or geologist (Fig. 3). However, boundaries can also be evaluated as the interface positions that have high mean information entropy values in 2D (Shi and Wang 2021). This study expands the estimation of the boundaries throughout the study area in both 2D and 3D to avoid crude approximations and assumptions associated with linear interpolation. The proposed hybrid approach will identify nonlinear boundaries separating geomaterials and better describe the true subsurface stratigraphy for the study site.

To give an initial description of the boundaries between geomaterial layers in the surface area, two-dimension (2D) images in the longitudinal and transverse axes are displayed in Fig. 11. Boundaries obtained from tracing gridded spatial positions with high location mean information entropy are displayed with the black bold line while those from the linear interpolated method are displayed with the black dashed line. The boundaries have a continuous bold line in the longitudinal direction with one geomaterial category placed on another. In the trans-

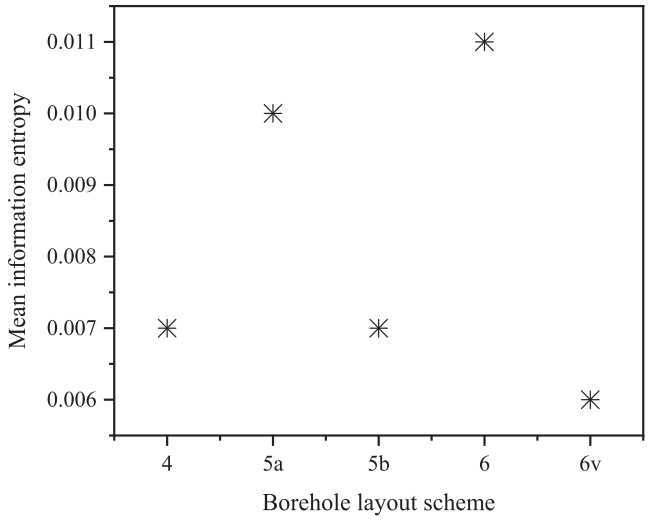

Fig. 10. Overall mean information entropy for various borehole layout schemes (Ng et al., 2023).

verse direction, however, the sand and siltstone layers seem to be the only geomaterials spanning throughout the length of this direction. There are noticeable differences in the boundaries obtained from both the proposed and linear interpolated methods. For example, the boundaries between the sand layer and the sand with silt and gravel in the longitudinal direction had a difference ranging from 0.3 to 3 m as shown in Fig. 11(a). Such differences could lead to a biased prediction of geomaterial properties, expected pile penetration length, and pile resistance estimation.

A 3D boundary description showing the geomaterial layers with a H-pile driven to a penetration depth of 16.8 m is shown in Fig. 12. The linear interpolation method considering site-specific geological history which is the current practice is restricted to 2D. It is nearly impossible to determine the layer boundaries of sparse data with linear interpolation in 3D. The 3D image gives a vivid and realistic arrangement of the different geomaterial layers, unlike the 2D image provided the geological profile is accurately predicted. For example, the sand layer did not span throughout the surface area of the spatial study site in 3D (Fig. 12) unlike how it did throughout the length of the 2D longitudinal and transverse direction (Fig. 11). The estimated thickness of each geomaterial layer on the shaft of the 16.8 m driven pile is 4.5 m, 9 m, and 3.3 m for sand with silt and gravel, sand, and sandy siltstone, respectively. These geomaterial layer depths are important for the reliable design of pile foundations. Table 5 presents the results comparing geomaterial layer thickness for a

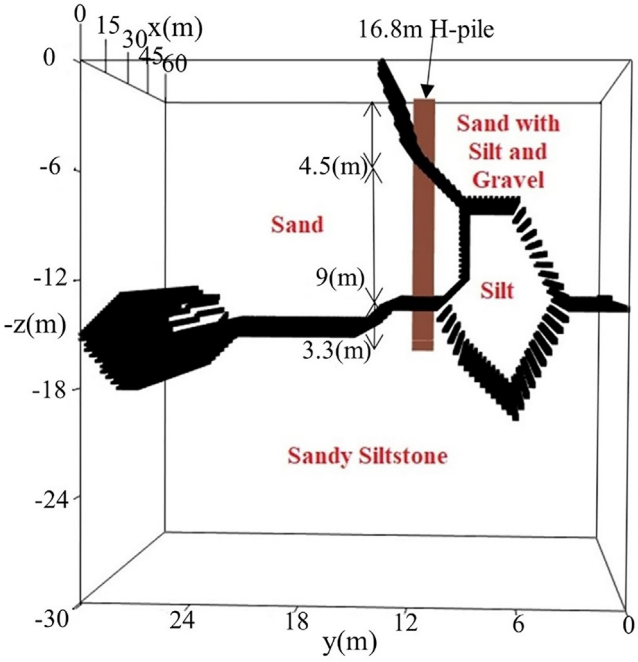

Fig. 12. 3D boundary descriptions for geomaterials and 16.8 m pile penetration (Ng et al., 2023).

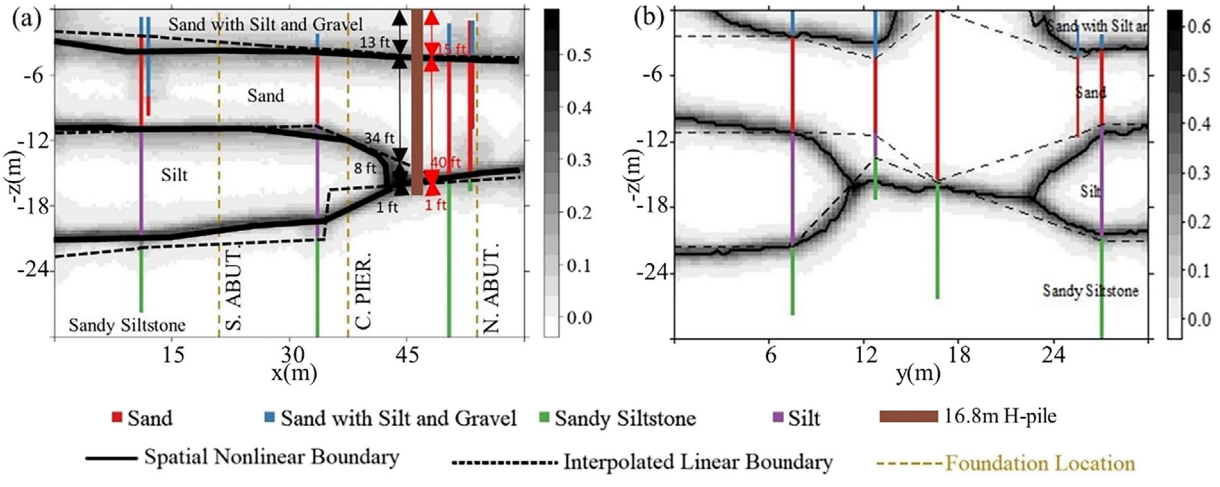

Fig. 11. 2D boundary descriptions for geomaterials in the (a) longitudinal direction and (b) transverse direction (Ng et al., 2023).

Table 5
Geomaterial layers depth for a 16.8 m pile penetration as determined from the boundaries (Ng et al., 2023).

Boundary determination method	Geomaterial layer thickness (m)			
	Sand with Silt and Gravel	Sand	Silt	Sandy Siltstone
2D linear interpolation	3.9	10.2	2.4	0.3
2D proposed hybrid	4.5	12	0	0.3
3D proposed hybrid	4.5	9	0	3.3
Closest borehole to the pile	4.5	10.8	0	1.5

16.8 m pile penetration as determined from the boundary layers using the 2D and 3D proposed hybrid approach and the 2D restricted linear interpolation method considering site-specific geological history. The results showed discrepancies in the different methods of estimating the depth of the geomaterial layer along the driven pile from the head to the toe. There is no borehole at the exact position of the driven pile to validate the results. However, extensive scrutiny of Fig. 11(a) showed silt to be absent in the geomaterial layers revealed by the closest borehole to the pile. A similar pattern is shown in the geomaterial layers revealed by the proposed hybrid approach (Table 5). Hence, the proposed hybrid approach of using spatial boundaries to determine geomaterial layers is an alternative recommendation for practice.

5. Conclusions

The geomaterial subsurface stratigraphy can play an important role in the design of reliable pile foundations, particularly in an IGM, which has variable characteristics and properties. The paper presents an innovative and comprehensive hybrid approach for quantifying the uncertainty associated with modeling geomaterial subsurface profiles from a site investigation. The approach includes the estimation, simulation, and reduction of geological uncertainty as well as the determination of the geomaterial layer boundaries. A hybrid approach consisting of spMC methodology and Monte Carlo simulation is used to accommodate the sparse data in a 3D layout. This extends the current engineering practice of using linear interpolations to draw layer boundaries between observed geomaterials to the usage of geostatistically determined nonlinear spatial boundaries. It also provides the background required for the unified analysis of geological and property uncertainties to design a reliable bridge foundation system.

The proposed hybrid approach is applied to the borehole data from the Lodgepole Creek bridge replacement project in Wyoming. Estimation of transition rate matrices is done with the Maximum Entropy Method. Predictions are performed using Indicator Kriging (IK), Indicator CoKriging (CK), and Multinomial Categorical Prediction (MCP) methods, and geological uncertainty is propagated using Monte Carlo simulation. Mean information entropy is used to quantify the geological uncertainty associated with the predicted geomaterial categories. The following conclusions are drawn from the results of this study:

1) The transition rate matrices and transition probability matrices can be estimated in 3D using spMC models with sparse borehole data.
2) The MCP method performs the best in minimizing overall mean information entropy for the prediction algorithms in this case study. The cross-validation of the multiple simulations indicates that the geomaterial predictions are generally accurate but depend upon the consistency of the site-specific borehole data.
3) The hybrid approach allows for the evaluation of borehole schemes to reduce geological uncertainty for all geomaterials or a particular geomaterial. This in turn facilitates the establishment of an optimum SI plan for bridge pile foundation design and construction.
4) The estimation of the geomaterial layer boundary is critical to the design of an adequate bridge foundation system. The hybrid approach provides a means by which to compare the current engineering practice of linear interpolation to obtain spatial boundaries between geomaterial layers. The use of geostatistical boundaries will improve the estimation of pile penetration length and resistances during the pile design stage.
5) Reducing the geological uncertainties and accurate determination of spatial geomaterial boundaries will improve the design reliability and safety of bridge foundation systems following the Load and Resistance Factor Design philosophy.

6. Copyright statement

tions of Driven Piles on Intermediate Geomaterials, TPF 5–391.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their gratitude to the research support from 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 Montana Department of Transportation under the Grant RS05219. Additional support from Mountain Plains Consortium is also appreciated.

References

Allard, D., D'Or, D., Froidevaux, R., 2011. An efficient maximum entropy approach for categorical variable prediction. Eur. J. Soil Sci. 62, 381–393. https://doi.org/10.1111/j.1365-2389.2011.01362.x.

Ampomah, W., Balch, R.S., Cather, M., Will, R., Gunda, D., Dai, Z., Soltanian, M.R., 2017. Optimum design of CO2 storage and oil recovery under geological uncertainty. Appl. Energy 195, 80–92. https://doi.org/10.1016/j.apenergy.2017.03.017.

Benndorf, J., Dimitrakopoulos, R., 2013. Stochastic long-term production scheduling of iron ore deposits: integrating joint multi-element geological uncertainty. J. Min. Sci. 49, 68–81. https://doi.org/10.1134/S1062739149010097.

Bivand, R.S., Pebesma, E.J., Virgilio, G.-R., 2013. Applied spatial data analysis with R, Second. ed. Springer, New York.

Bogaert, P., 2002. Spatial prediction of categorical variables: the Bayesian maximum entropy approach. Stoch. Environ. Res. Risk Assess. 16, 425–448. https://doi.org/10.1007/s00477-002-0114-4.

Carle, S.F., Fogg, G.E., 1996. Transition probability-based indicator geostatistics. Math. Geol. 28, 453–476. https://doi.org/10.1007/bf02083656.

Carle, S.F., Fogg, G.E., 1997. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Math. Geol. 29, 891–918. https://doi.org/10.1023/A:1022303706942.

de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. Hydrogeol. J. 13, 161–183. https://doi.org/10.1007/s10040-004-0432-3.

Deng, Z.P., Li, D.Q., Qi, X.H., Cao, Z.J., Phoon, K.K., 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.

Deng, Z.P., Jiang, S.H., Niu, J.T., Pan, M., Liu, L.L., 2020. Stratigraphic uncertainty characterization using generalized coupled Markov chain. Bull. Eng. Geol. Environ. 79, 5061–5078. https://doi.org/10.1007/s10064-020-01883-y.

Elfeki, A.M., 2006. Prediction of contaminant plumes (shapes, spatial moments and macrodispersion) in aquifers with insufficient geological information. J. Hydraul. Res. 44, 841–856. https://doi.org/10.1080/00221686.2006.9521735.

Elfeki, A.M.M., Dekking, F.M., 2007. Reducing geological uncertainty by conditioning on boreholes: the coupled Markov chain approach. Hydrogeol. J. 15, 1439–1455. https://doi.org/10.1007/s10040-007-0193-x.

Feng, R., Luthi, S.M., Gisolf, D., 2018. Simulating reservoir lithologies by an actively conditioned Markov chain model. J. Geophys. Eng. 15, 800–815. https://doi.org/10.1088/1742-2140/aaa0ff.

Gao, R.S., Sun, A.Y., Nicot, J.P., 2016. Identification of a representative dataset for long-term monitoring at the Weyburn CO2-injection enhanced oil recovery site, Saskatchewan. Canada. Int. J. Greenh. Gas Control 54, 454–465. https://doi.org/10.1016/j.ijggc.2016.05.028.

Goovaerts, P., 1994. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. Math. Geol. 26, 389–411. https://doi.org/10.1007/BF02089230.

Grabski, F., 2014. Semi-Markov Processes: Applications in System Reliability and Maintenance. Elsevier.

Huffman, J.C., Strahler, A.W., Stuedlein, A.W., 2015. Reliability-based serviceability limit state design for immediate settlement of spread footings on clay. Soils Found. 55, 798–812. https://doi.org/10.1016/j.sandf.2015.06.012.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2015. An introduction to statistical learning: with applications in R. Springer, New York.

Kutner, M.H., Nachstsheim, C.J., Neter, J., 2004. Applied linear regression models, Fourth. ed. McGraw Hill, Boston.

Li, W., 2007. A fixed-path Markov chain algorithm for conditional simulation of discrete spatial variables. Math. Geol. 39, 159–176. https://doi.org/10.1007/s11004-006-9071-7.

Li, W., Zhang, C., 2015. A comment on Sartore's "spMC: Modelling spatial random fields with continuous lag Markov chains.".

Li, Z., Wang, X., Wang, H., Liang, R.Y., 2016. Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. Eng. Geol. 201, 106–122. https://doi.org/10.1016/j.enggeo.2015.12.017.

Li, W., Zhang, C., 2007. A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. Soil Sci. Soc. Am. J. 71, 656–668. https://doi.org/10.2136/sssaj2006.0173.

Li, W., Zhang, C., 2010. Linear interpolation and joint model fitting of experimental transiograms for Markov chain simulation of categorical spatial variables. Int. J. Geogr. Inf. Sci. 24, 821–839. https://doi.org/10.1080/13658810903127991.

Lindsay, M.D., Aillères, L., Jessell, M.W., de Kemp, E.A., Betts, P.G., 2012. Locating and quantifying geological uncertainty in three-dimensional models: analysis of the Gippsland Basin, southeastern Australia. Tectonophysics 546–547, 10–27. https://doi.org/10.1016/j.tecto.2012.04.007.

Montiel, L., Dimitrakopoulos, R., Kawahata, K., 2016. Globally optimising open-pit and underground mining operations under geological uncertainty. Min. Technol. 125, 2–14. https://doi.org/10.1179/1743286315Y.0000000027.

Myers, D.E., 1982. Matrix formulation of co-kriging. J. Int. Assoc. Math. Geol. 14, 249–257. https://doi.org/10.1007/BF01032887.

Ng, K.W., Masud, N.B., Oluwatuyi, O.E., Wulff, S.S., 2023. Comprehensive field test and geotechnical investigation program for development of LRFD recommendations of driven piles on intermediate geomaterials. Cheyenne, WY.

Oluwatuyi, O.E., Holt, R., Rajapakshage, R., Wulff, S.S., Ng, K.W., 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. https://doi.org/10.1080/17499518.2022.2046783.

Oluwatuyi, O.E., Rajapakshage, R., Wulff, S.S., Ng, K.W., 2022b. Quantifying geological uncertainty using conditioned spatial Markov chains. Geo-Congress 2022, 436–445. https://doi.org/10.1061/9780784484036.043.

Park, E., 2010. A multidimensional, generalized coupled Markov chain model for surface and subsurface characterization. Water Resour. Res. 46, 1–15. https://doi.org/10.1029/2009WR008355.

Park, E., Elfeki, A.M.M., Song, Y., Kim, K., 2007. Generalized Coupled Markov Chain Model for Characterizing Categorical Variables in Soil

Mapping. Soil Sci. Soc. Am. J. 71, 909–917. https://doi.org/10.2136/sssaj2005.0386.

Pedretti, D., 2020. Heterogeneity-controlled uncertain optimization of pump-and-treat systems explained through geological entropy. GEM - Int. J. Geomath. 11. https://doi.org/10.1007/s13137-020-00158-8.

Prospere, K., McLaren, K., Wilson, B., 2016. Substrate mapping of three rivers in a Ramsar wetland in Jamaica: a comparison of data collection (hydroacoustic v. grab samples), classification and kriging methods. Mar. Freshw. Res. 67, 1771–1795. https://doi.org/10.1071/MF15033.

Qi, X.H., Li, D.Q., Cao, Z.J., Tang, X.S., 2016. Uncertainty analysis of slope stability considering geological uncertainty, in: Proceedings of 12th International Symposium on Landslides and Engineered Slopes–Experience, Theory and Practice. pp. 1685–1693.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing

Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Adv. Water Resour. 36, 36–50. https://doi.org/10.1016/j.advwatres.2011.04.006.

Sartore, L., 2013. SpMC: Modelling spatial random fields with continuous Lag Markov chains. R J. 5, 16–28. https://doi.org/10.32614/rj-2013-022.

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.

Sartore, L., 2019. spMC-package: Continuous Lag Spatial Markov Chains.

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, 1669–1689. https://doi.org/10.1139/cgj-2020-0327.

Tran, T.T., Han, S.R., Kim, D., 2018. Effect of probabilistic variation in soil properties and profile of site response. Soils Found. 58, 1339–1349. https://doi.org/10.1016/j.sandf.2018.07.006.

Wang, Y., Cao, Z., 2013. Expanded reliability-based design of piles in spatially variable soil using efficient Monte Carlo simulations. Soils Found. 53, 820–834. https://doi.org/10.1016/j.sandf.2013.10.002.

Wang, X., Wang, H., Liang, R.Y., Liu, Y., 2019. A semi-supervised clustering-based approach for stratification identification using borehole and cone penetration test data. Eng. Geol. 248, 102–116. https://doi.org/10.1016/j.enggeo.2018.11.014.

Wellmann, J.F., 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, Q., Zhu, H., 2018. Collaborative 3D geological modeling analysis based on multi-source data standard. Eng. Geol. 246, 233–244. https://doi.org/10.1016/j.enggeo.2018.10.001.