Model Updating for Geotechnical Design and Assessment

Cotoarbă Dafydd, Daniel Straub, Ian F.C. Smith
Technical University of Munich, Germany
dafydd.cotoarba@tum.de

Abstract. A categorization of subsoil characterization approaches is used as the basis for evaluating them according to their capacity for model updating, quantifying uncertainties and modelling complex geometries. The information of using the right model at the right time is one of the most important sources of engineering knowledge. To ensure an equitable comparison under comparable starting conditions, three synthetic soil topologies are generated, each featuring relevant geological processes for subsoil characterization. In this study, we evaluated the performance of a voxel-based approach for subsoil characterization using the discrete output of Gaussian Process Regression (GPR). The evaluation based on the three metrics shows that the potential advantages are simplicity, ability to integrate new data and quantification of prediction uncertainty. Future work will focus on analysing additional approaches, such that an appropriate framework for digital-twins for geotechnical design and assessment can be defined.

1. Introduction

Geotechnical engineering applies the principles of soil mechanics to design structures in soil. In order to achieve this, engineers must study site-specific soil properties and their deformation behaviour under loads. Das (2021) defines soil mechanics as the study of soil to understand the physical properties and their behaviour when subjected to load stresses. Lack of understanding of site-specific soil mechanics can negatively impact design decisions and lead to high economic consequences and safety issues, such as project delays and structural failure. However, characterizing soils and their behaviour is a complex task that relies on data from multiple sources, such as boreholes, cone penetration tests (CPTs), trenches, on-site (geophysical) tests, and site-specific samples tested in laboratories (Zhang et al. 2018).

To extract information from measurement data, several interpretation methods have been developed using data such as borehole logs, geological maps and cross-sections. Technological advancements have led to a rise in the amount of geotechnical data collected during a project (Chandler, 2011; Phoon, 2019). Advanced information management tools are required to manage and integrate such diverse data in real-time to be able to offer support for decision making during construction (Zhou, Ding and Chen, 2013). The use of building information modelling (BIM) has gained popularity in the geotechnical community due to their ability to structure project-related data and attribute it to individual components of 3D soil topologies (Tawelian and Mickovski, 2016; Zhang *et al.*, 2018). However, while the traditional BIM approach is useful for information management, it lacks the ability to quantify uncertainties. This is because traditional BIM approaches assume full knowledge of reality, while geotechnical engineering involves obtaining a subsoil topology by analysing a finite number of geological survey points (Wu *et al.*, 2021). As a result, there may be significant uncertainty associated with site-specific knowledge of geotechnical properties.

The assumption of the probabilistic nature of soils is a result of the complex geological formation process of soils and our incomplete knowledge (Hu *et al.*, 2020). Although the physical causes of soil formation are deterministic and obey the laws of physics, it is currently impossible to fully understand how they combine. Furthermore, it is difficult to study their variation over time with incomplete knowledge. As a result, soil formation is assumed to be random (Webster, 2000). For the task of estimating geotechnical properties, Kulhawy et al. (2006) refer to this uncertainty as the inherent soil variability. It introduces knowledge uncertainty, as full knowledge of all properties of the subsoil is not possible, making it difficult to quantify the accuracy of a characterization.

In addition to the uncertainty related to soil formation, there are three other principal sources of uncertainty in geotechnical engineering: measurement errors, transformation uncertainties and statistical uncertainties. Measurement errors occur during the data collection process, which involves using specific equipment, procedures, and personnel. Transformation uncertainties occur when measurements are transformed into soil properties. Statistical uncertainty is the result of the limited availability of in-situ measurements, which is a common issue. Sparse data may result in unreliable soil descriptions that are overfitted to local data and underestimate uncertainties (Pyrcz and Deutsch, 2014).

The uncertainties described above make subsoil characterization particularly challenging. An ideal approach should function with sparse data, allow real-time model updating, and enable the explicit quantification of knowledge of uncertainty. Geostatistical methods have been developed to address these challenges by modelling the distribution of soil types and their properties. These methods help integrate information from multiple data sources to infer representative statistics (Pyrcz and Deutsch, 2014). The effectiveness of geostatistical methods depends on the suitability of their application. This is important engineering knowledge that is rarely made explicit.

In this paper we propose a classification of subsoil characterization approaches to facilitate an analysis of their strengths and limitations. To ensure an equitable comparison with similar starting conditions, we have generated three synthetic soil topologies, each featuring relevant geological processes for subsoil characterization. By using only simulated measurements to recreate the

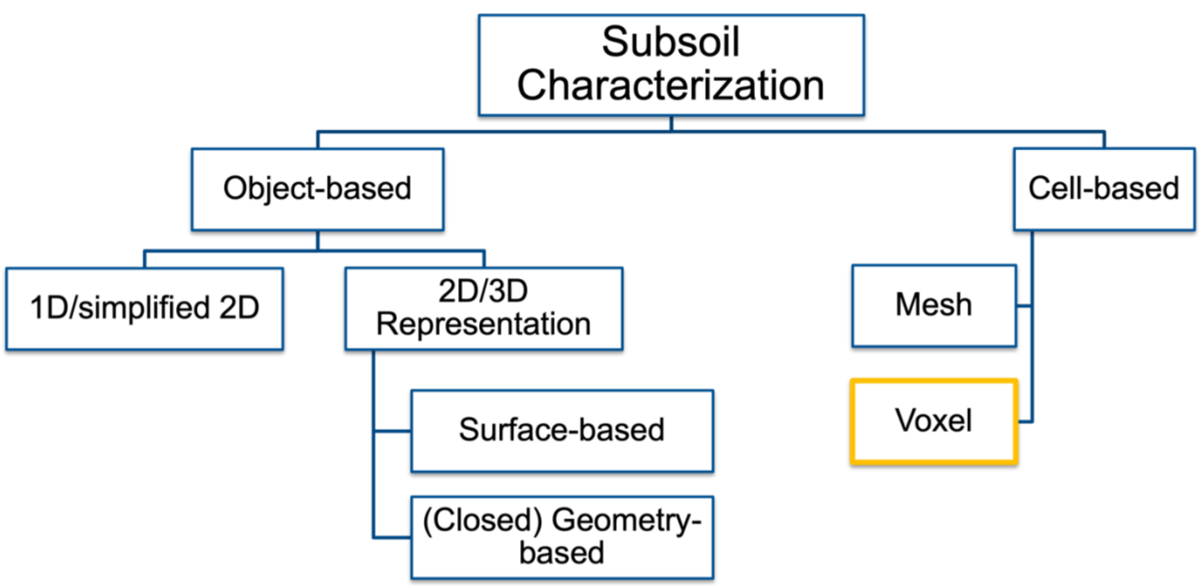

Figure 1 - Categorization for subsoil characterization approaches. The approach that is used in this paper is highlighted.

ground truth, subsoil characterization methods can be evaluated on metrics such as accuracy for complex geometries, data updating capabilities, and uncertainty quantification. Our study focuses on a voxel, cell-based subsoil characterization approach, serving as a proof-of-concept for this methodology, which can now be extended to analyse additional approaches.

2. Classification of probabilistic subsoil characterization methods

Figure 1 presents a classification of subsoil characterization approaches. At the highest level categorizes approaches into two types according to their representation: object-based and cell-based.

2.1 Object-based approach

The object-based approach involves a representation of the geometry of geological bodies and the relationship between layers directly (Lyu et al. 2021).

The 1D approach is utilized to identify stratigraphic layers from in-situ measurements assuming constant layers throughout the construction site. For example, Cao and Wang (2013) present a Bayesian approach to identifying soil layers and their thickness through CPT measurements, which differs from deterministic approaches that imply correlations based on empirical by clustering similar data points to identify layers. The Bayesian approach allows for the quantification of uncertainties. The approach output can serve as input for 2D/3D subsoil characterization methods.

The object-based approach for the 3D representation often differs in the definition of the function used to describe the geological interfaces. The surface-based approach functions, which describe open shapes, are conditioned to geological data. The form of the function influences the interpolation for unknown points. After interpolation, expert knowledge is used to ensure coherence of the geological interfaces from a soil mechanical point of view. It is still common to perform this type of correction manually (Lyu et al. 2021). However, current work is focused on formalizing expert knowledge to provide greater support (Zhu et al. 2012).

In the geometry-based approach, closed shapes (e.g. cylinder, circles, rectangles) are used to describe geological interfaces. However, they offer limited flexibility as they are difficult to condition to experimental data. They are typically used for reservoir modelling, where the focus lies on estimating the volume of for example oil pockets and where the exact trajectory of geological interfaces is less relevant (Pyrcz and Deutsch, 2014).

2.2 Cell-based approach

The second approach to subsoil characterization is the cell-based approach, in which the subsoil is discretized into a grid of cells, where each is described by a set of parameters and soil type. We distinguish between two possible formats of the approach.

First, the mesh-based approach, unknown points are calculated using weighted averages of observational data. The chosen mesh structure will influence the boundaries between different soil layers. The calculation process is efficient, conforms to geological laws and respect the observations. However, it has limitations when complex geological interfaces, such as overturned folds, are modelled (Lyu *et al.*, 2021).

A sub-case of this approach, the voxel-based uses a structured grid to characterize subsoils. It is commonly used for stochastic modelling approaches when soil properties are described as random variables and spatial variability is modelled as a random field (Wang *et al.*, 2018). Investigations are usually performed on the discretized random field. As highlighted in Figure 1, in this work a voxel, cell-based approach is pursued.

3. Numerical Investigation

Figure 2: Topology 1 showing the process of sequential deposition of soil layers.

Figure 3: Topology 2 showing the erosion process in combination with sequential deposition of soil layers.

Figure 2: Topology 3 is a result of simultaneous deposition of layers including lenses.

To compare subsoil characterization approaches, we have created three synthetic 2D soil topologies, as illustrated in Figures 2-4. Each topology was designed to represent at least one geological process. Topology 1 represents the process of sequential deposition of soil layers, while Topology 2 incorporates the effect of erosion during deposition, resulting in the next layer being deposited in the space resulting from erosion. Topology 3 represents a simultaneous deposition process, resulting in the formation of so-called lenses, which are islands of one soil type surrounded by another. For the description of the soil layers the common Soil Behaviour Index (SBT) I_C is used, which is defined in Equation 1, below.

A commonly used method for in-situ soil investigation is Cone Penetration Testing (CPT), which measures the mechanical response of soil to penetration using a cylindrical steel probe. This testing method provides continuous data over depth and is preferred over other methods due to its speed, repeatability, and affordability (Robertson, 2009). The data that is obtained from CPT includes the cone tip resistance q_T and sleeve friction f_c (Mayne, Christopher and DeJong, 2002).

To synthetically generate CPT data, transformation functions are used to populate the synthetic topologies with q_T and f_s data. For each soil type, a Gaussian random field is generated for each property. To describe the random field, we use the approximate guidelines for inherent soil variability that are described in Phoon and Kulhawy (1999). The parameter values that have been employed are provided in Table 1.

Table 1: Summary of soil parameters for the description of the random fields with μ_x mean of property x, Σ_x covariance for x and l_x, l_y scale of fluctuation in x and y direction.

I_C	Soil Type	μ_{q_T} $[MPa]$	Σ_{q_T} $[MPa]$	$(l_{x_{q_T}}, l_{y_{q_T}})$ $([m],[m])$	μ_{f_s} $[MPa]$	Σ_{f_s} $[MPa]$	$(l_{x_{fs}}, l_{y_{fs}})$ $([m],[m])$
2	Clays	0.5	0.2	$(30, 0.5)$	0.035	0.2	$(30, 0.5)$
4	Silt mixtures	2	0.2	$(40, 1)$	0.08	0.2	$(40, 1)$
5	Sand mixtures	10	0.3	$(40, 1.5)$	0.25	0.3	$(40, 1.5)$
6	Sands	15	0.2	$(40, 2)$	0.11	0.2	$(40, 2)$

This results in a random field description for each of the three soil topologies. One sample for each topology is drawn to obtain a ground truth topology for the numerical investigation. They can be seen in Figures 2-4.

Based on empirical data, Phoon (1995) states that for CPT measurements an error of $\sim 35\%$ can be expected. As a result, an error term ϵ_4 that accounts for measurement errors is added to each simulated CPT measurement. ϵ_4 is normally distributed with zero mean with standard deviation equal to 0.35.

The simulated measurements of q_T and f_s are used as an input in Equations (1) and (2) to obtain the soil behaviour type (SBT) index I_c defined in Robertson and Wride (1998) and adapted in Robertson (2009):

$$I_c = \sqrt{\left(3.47 - \log_{10} \frac{q_T}{P_a}\right)^2 + (\log_{10} F_R + 1.22)^2} \tag{1}$$

where P_a is the atmospheric pressure (approx. 100 kPa) and F_R the normalized friction ratio:

$$F_R = \frac{f_s}{q_T} * 100\% \tag{2}$$

Depending on the value I_c that is obtained for given measurements, the soil can be categorized in one of six SBTs (see Table 2).

Table 2: Classification of soil according to the soil behaviour type index by Robertson (2009)

Zone	**Soil Behavior Type**	I_c
2	Organic Soils: clay	> 3.60

3	Clays: silty clay to clay	$2.95 - 3.60$
4	Silt mixtures: clayey silt to silty clay	$2.60 - 2.95$
5	Sand mixtures: silty sand to sandy silt	$2.05 - 2.60$
6	Sands: clean sand to silty sand	$1.31 - 2.05$
7	Gravelly sand to dense sand	< 1.31

4. Gaussian Process Regression

A widely used cell-based approach to subsoil characterization, Gaussian Process Regression (GPR), is investigated (Rasmussen and Williams, 2005). In geostatistics, it is commonly also referred to as Kriging. It is a regression analysis tool that describes a conditional random field by combining a prior multinormal distribution with observation data to obtain a posterior distribution. The posterior distribution is used to describe the statistics at unknown points.

In this work GPR is used to obtain I_C based on the simulated CPT measurements. A grid of $100x100$ cells is defined. To describe the correlation between two points of the grid in relation to their relative distance, the anisotropic common squared exponential function is used, see Equation (3).

$$\Sigma_{\mathrm{x}}(a, b) = \exp\left(\frac{d(a_x, b_x)^2}{2l_X^2}\right) \tag{3}$$

where $d(a, b)$ is the distance between two points and l_x is the correlation length in the x-direction. An optimization is performed to obtain the correlation length (Pedregosa *et al.*, 2011).

The output of the GPR is a probability distribution for each soil type at each cell, which can be used for classification tasks such as identifying soil layers. The accuracy of the resulting estimates is dependent on the quality and amount of available data. In geotechnical engineering, the number of measurements is often limited due to cost considerations. Therefore, it is of interest to investigate the effect of varying amount of data on the uncertainty of the subsoil characterization and to investigate how to best represent and communication this uncertainty.

4.1 Confidence

The advantage of utilizing Gaussian processes for classification is their ability to quantify uncertainties associated with classifications. For each cell, the Gaussian process returns a probability for each possible soil type. The results of this approach can be seen in Figures 5a-f. Let us consider Topology 1 with six CPT measurements (see Figure 5a). An initial visual comparison shows good agreement to the ground truth (Figure 2). Figure 5b shows the confidence in the classification. Two conclusions can be drawn from this plot: (1) there is high uncertainty at the boundaries between two soil layers, but less within a single soil layer, and (2) there is high uncertainty at locations where I_C obtained from the CPT measurements is faulty. This suggest that at these locations, the confidence in the classification is low, as multiple layers could potentially represent the data. In future work, the confidence values obtained from the Gaussian process can be used, for example, to guide the selection of optimal locations for performing the next observations.

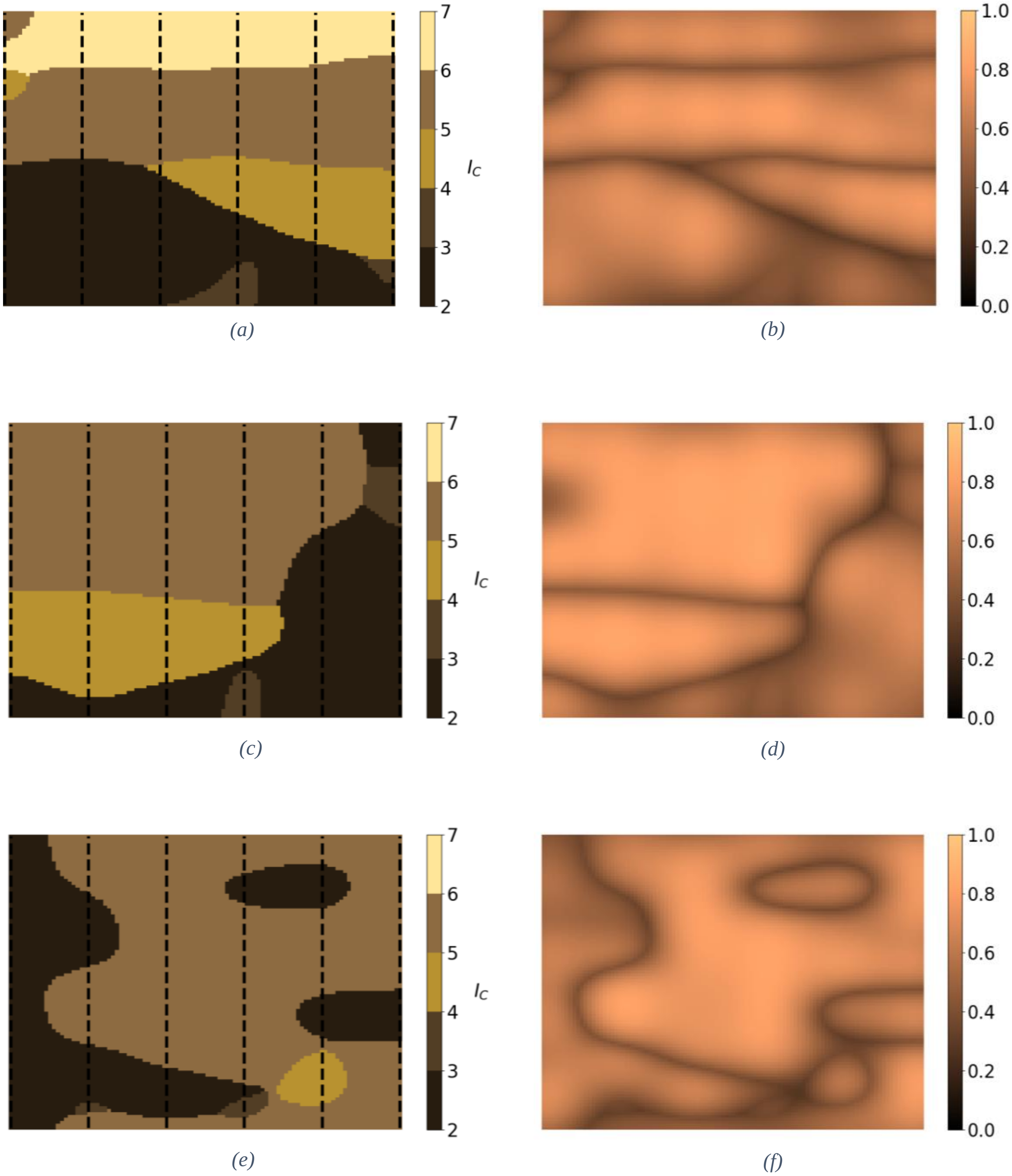

Figure 5: Left column: Resulting distribution of I_C for the Topologies and six CPT measurements (that are visualized with the black dotted lines); Right column: Display of the probability for identified I_C, which has the highest probability among all possible soil layers, at each grid point

4.2 Accuracy

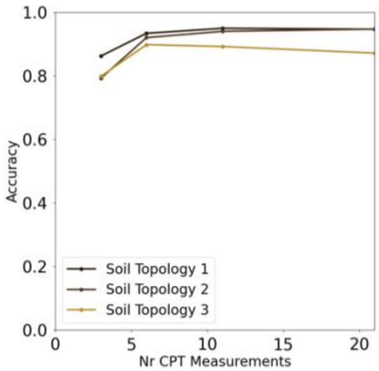

Figure 3: Accuracy of characterization approach for different number of CPT measurements

Table 3: Accuracy of the characterization approach for different number of CPT measurements

Soil Topology	**Number of CPT measurements**			
	3	**6**	**11**	**21**
1	0.91	0.96	0.96	0.96
2	0.87	0.95	0.95	0.96
3	0.80	0.92	0.91	0.91

To measure the performance of the equation the commonly used accuracy metric is used, as described in Equation (4).

$$Accuracy = \frac{\#Correct\ Predictions}{\#Total\ Predictions}. \tag{4}$$

The number of correctly predicted cells is divided by the total number of cells.

As shown in Table 2, the accuracy of the classification improves with an increasing number of CPT measurements. For an equidistant grid the improvement follows what seems a logarithmic path, with a high initial improvement that reaches a plateau (see Figure 11).

This finding is relevant for data-driven inspection strategies as it quantifies the value of additional information, indicating a point at which the marginal benefit of additional measurements diminishes. In the following analysis, each case will be considered individually.

The approach works well for Topology 1 where sequential deposition of different layers is considered. A visual inspection of Figure 5a shows that the largest source of error is the misclassification of the I_{SBT} from the CPT data.

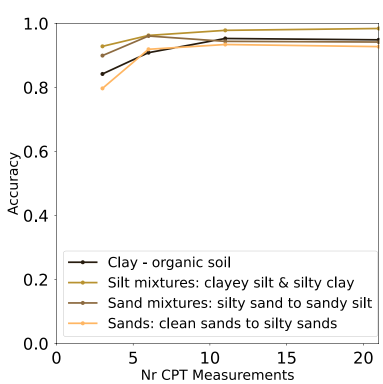

Figure 12: Accuracy for each soil type for Topology 1

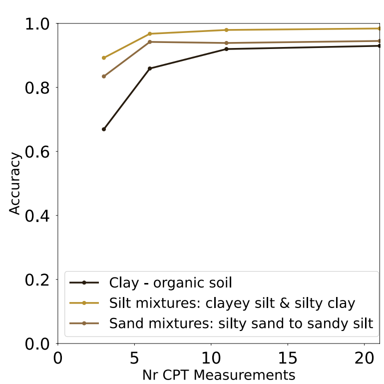

Figure 13: Accuracy for each soil type for Topology 2

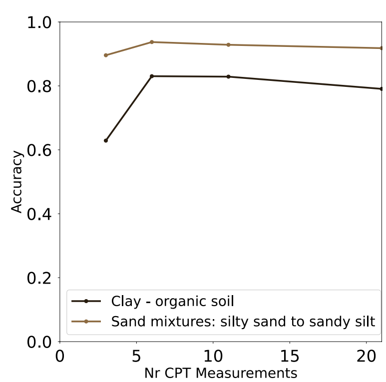

Figure 4: Accuracy for each soil type for Topology 3

Erosion, as presented in Topology 2 is more challenging for this approach. If the accuracy of each soil layer is considered individually (Figure 13), the accuracy varies largely between them. The largest source of error is the erosion in the clay layer (see Figure 5c). This is due to the correlation function, which offers equal importance to each CPT measurement.

The complexity of the layer distribution resulted from the simultaneous deposition is a large uncertainty for Topology 3. To obtain a more detailed representation, more measurements at critical points are required. In future work, it is of interest to research methods for the identification of such critical points and to quantify whether or not additional information will reduce the uncertainties. In addition, quantifying the value of additional observations can support the decision-making process to determine the best next step for investigation (Jiang, Papaioannou and Straub, 2020).

5. Conclusion

Our study has demonstrated potential advantages of categorizing subsoil characterization approaches based on their representation type for identifying their strengths and weaknesses. To compare various approaches, we generated three synthetic soil topologies, each featuring relevant geological processes. We evaluated the performance of a voxel-based approach for subsoil characterization on its capacity for data updating, uncertainty quantification, and modelling of complex geometries. The potential advantages of GPR are its simplicity, ability to integrate new data, and quantify prediction uncertainties. Future work will focus on analysing other approaches to improve our explicit understanding of the knowledge requirements for a digital-twin strategy for geotechnical design and assessment.

6. References

Cao, Z. and Wang, Y. (2013) 'Bayesian Approach for Probabilistic Site Characterization Using Cone Penetration Tests', *Journal of Geotechnical and Geoenvironmental Engineering*, 139(2), pp. 267–276. Available at: https://doi.org/10.1061/(asce)gt.1943-5606.0000765.

Chandler, R.J. (2011) 'Geotechnical Data Transfer and Management for Large Construction Projects and National Archives', in.

Das, B.M. (2021) *Principles of geotechnical engineering*. Cengage learning.

Hu, Y. *et al.* (2020) 'Bayesian Supervised Learning of Site-Specific Geotechnical Spatial Variability from Sparse Measurements', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 6(2). Available at: https://doi.org/10.1061/ajrua6.0001059.

Jiang, S.-H., Papaioannou, I. and Straub, D. (2020) 'Optimization of Site-Exploration Programs for Slope-Reliability Assessment', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 6(1), p. 04020004. Available at: https://doi.org/10.1061/AJRUA6.0001042.

Kok-Kwang, P. (1995) *Reliability-based design of foundations for transmission line structures*. Cornell University. Available at: https://www.proquest.com/openview/da0e0da72baa27c58382223e243b3948/1?pq-origsite=gscholar&cbl=18750&diss=y (Accessed: 27 March 2023).

Kulhawy, F.H. *et al.* (2006) 'Reliability-based design of foundations for transmission line structures', *Electrical Transmission Line and Substation Structures: Structural Reliability in a*

Changing World - Proceedings of the 2006 Electrical Transmission Conference, 218, pp. 184–194. Available at: https://doi.org/10.1061/40790(218)17.

Lyu, M. *et al.* (2021) ‘A parametric 3D geological modeling method considering stratigraphic interface topology optimization and coding expert knowledge’, *Engineering Geology*, 293. Available at: https://doi.org/10.1016/j.enggeo.2021.106300.

Mayne, P., Christopher, B. and DeJong, J. (2002) ‘Subsurface Investigations-- Geotechnical Site Characterization: Reference Manual’.

Pedregosa, F. *et al.* (2011) ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, 12(85), pp. 2825–2830. Available at: http://jmlr.org/papers/v12/pedregosa11a.html.

Phoon, K.K. (2019) ‘The story of statistics in geotechnical engineering’, *https://doi.org/10.1080/17499518.2019.1700423*, 14(1), pp. 3–25. Available at: https://doi.org/10.1080/17499518.2019.1700423.

Phoon, K.-K. and Kulhawy, F.H. (1999) *Characterization of geotechnical variability*.

Pyrcz, M.J. and Deutsch, C. V (2014) *Geostatistical reservoir modeling*. Oxford university press.

Rasmussen, C.E. and Williams, C.K.I. (2005) ‘Gaussian Processes for Machine Learning’, *Gaussian Processes for Machine Learning* [Preprint]. Available at: https://doi.org/10.7551/MITPRESS/3206.001.0001.

Robertson, P.K. (2009) ‘Interpretation of cone penetration tests — a unified approach’, *https://doi.org/10.1139/T09-065*, 46(11), pp. 1337–1355. Available at: https://doi.org/10.1139/T09-065.

Robertson, P.K. and Wride, C.E. (2011) ‘Evaluating cyclic liquefaction potential using the cone penetration test’, *https://doi.org/10.1139/t98-017*, 35(3), pp. 442–459. Available at: https://doi.org/10.1139/T98-017.

Tawelian, L.R. and Mickovski, S.B. (2016) ‘The Implementation of Geotechnical Data into the BIM Process’, in *Procedia Engineering*. Elsevier Ltd, pp. 734–741. Available at: https://doi.org/10.1016/j.proeng.2016.06.115.

Wang, X. *et al.* (2018) ‘A hidden Markov random field model based approach for probabilistic site characterization using multiple cone penetration test data’, *Structural Safety*, 70, pp. 128–138. Available at: https://doi.org/10.1016/j.strusafe.2017.10.011.

Webster, R. (2000) *Is soil variation random?, Geoderma*. Available at: www.elsevier.nlrlocatergeoderma.

Wu, J. *et al.* (2021) ‘Development of Data Integration and Sharing for Geotechnical Engineering Information Modeling Based on IFC’, *Advances in Civil Engineering*, 2021. Available at: https://doi.org/10.1155/2021/8884864.

Zhang, J. *et al.* (2018) ‘The BIM-enabled geotechnical information management of a construction project’, *Computing*, 100(1), pp. 47–63. Available at: https://doi.org/10.1007/s00607-017-0571-8.

Zhou, Y., Ding, L.Y. and Chen, L.J. (2013) ‘Application of 4D visualization technology for safety management in metro construction’, *Automation in Construction*, 34, pp. 25–36.

Zhu, L. *et al.* (2012) ‘Building 3D solid models of sedimentary stratigraphic systems from borehole data: An automatic method and case studies’, *Engineering Geology*, 127, pp. 1–13. Available at: https://doi.org/10.1016/j.enggeo.2011.12.001.