Journal of GeoEngineering, Vol. 17, No. 4, pp. 231-247, December 2022
http://dx.doi.org/10.6310/jog.202212_17(4).6

ADDITIONAL OBSERVATIONS ON THE SITE RECOGNITION CHALLENGE

Kok-Kwang Phoon [1*], Jianye Ching [2]

ABSTRACT

One distinctive feature of geotechnical engineering is site uniqueness or site-specificity. However, there is no data-driven method to quantify site uniqueness. The corollary is that it is not possible to identify "similar" sites from big indirect data (BID) automatically and no method to combine sparse site-specific data with big indirect data to produce a quasi-site-specific model that is less biased compared to a generic model and less imprecise compared to a site-specific model. This "site recognition" challenge is difficult because site-specific data is MUSIC-X (Multivariate, Uncertain and Unique, Sparse, Incomplete, and potentially Corrupted with "X" denoting the spatial/temporal dimension). This paper presents the application of four data-driven methods (hybridization, hierarchical Bayesian model, record similarity method, site similarity method) to construct a quasi-site-specific transformation model between the undrained shear strength and normalized cone tip resistance. The similarity methods are "explainable", because a list of "similar" sites can be generated explicitly for inspection by the engineer. The effect of extrapolating the quasi-site-specific model beyond the range of the training dataset is also studied by comparing the performance of these models under routine validation (validation dataset is contained within the training dataset) and under external validation (validation dataset lies outside the training dataset). The hierarchical Bayesian model appears to be the best performing method thus far, but it suffers from a lack of "explainability". More research is needed to: (1) ascertain the number and/or type of soil properties needed to identify "similar" sites more robustly in the sense of producing more clustered results in existing soil classification charts (*e.g.,* Casagrande plasticity chart, Robertson CPT-based soil behavior type classification system) and/or producing the most accurate quasi-site-specific model, (2) understand the bias and precision of making inferences beyond the range of the training dataset, and (3) clarify the trade-off between explainability and inference (bias and precision).

Key words: Data-driven site characterization; transformation model; site recognition; explainability; generalizability.

1. INTRODUCTION

Site characterization is a cornerstone of geotechnical and rock engineering. "Data-driven site characterization" (DDSC) refers to any site characterization methodology that relies solely on measured data, both site-specific data collected for the current project and existing data of any type collected from past stages of the same project or past projects at the same site, neighboring sites, or beyond. Phoon *et al.* (2022a) outlined three challenges in DDSC: (1) ugly data, (2) site recognition, and (3) stratification. "Ugly" refers to the attributes of actual data that present significant difficulties to classical statistics. The most well-known attributes in geotechnical practice are sparsity (not enough data and/or measurements separated wide apart) and spatial variability (data are spatially correlated). Another well-known attribute is quality. Phoon (2020) argued that the jury is still out on many important questions pertaining to the *value of quality to decision making*: "Do we collect only high quality data that are limited in quantity or only lower quality data in larger quantity? Do we combine them? When shouldn't they be combined?"

The conventional wisdom is that statistics is not applicable based on the sparse attribute alone. Engineers commonly observed that data are insufficient for probabilistic analysis. For example, Schuppener (2011) opined "soil excavations and tests of the mechanical properties of soil never provide enough data to enable a probability calculation to be performed". This "curse of small sample size" is more conspicuous in geotechnical engineering than in structural engineering (Phoon 2017).

Site recognition is related to a fundamental feature in geotechnical practice, namely all sites are different to some extent (site specificity). For example, Clause 2.4.5.2(10) of Eurocode 7 (CEN 2004) "Characteristic values of geotechnical parameters" advises against pooling data from different sites into a single population for statistical analysis: "If statistical methods are employed in the selection of characteristic values for ground properties, such methods should differentiate between local and regional sampling and should allow the use of a priori knowledge of comparable ground properties".

Stratification requires an identification of soil type and a mapping of the spatial distribution of each soil type based on field tests typically conducted at discrete locations/depths in a soil mass. Geophysical data provide more continuous spatial coverage, but they are less common and not well studied (Sauvin *et al.* 2019). These challenges are not new. An engineer encounters them routinely in practice, but they are not regarded as challenges because judgment-based solutions are widely accepted as satisfactory. For instance, a single characteristic depth profile for a soil parameter is selected from actual spatially variable measurements by an engineer making a "cautious" judgment call. One expects different

Manuscript received September 6, 2022; revised October 23, 2022; accepted November 18, 2022.
[1*] Professor (corresponding author), Singapore University of Technology and Design, 8 Somapah Road 487372, Singapore (e-mail: kkphoon@sutd.edu.sg).
[2] Professor, Dept. of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan.

engineers to arrive at different characteristic profiles (Bond and Harris 2008). This is hardly surprising. An engineer applies experience gained from other sites to interpret limited data at the current site of interest. However, he/she will not regard all sites to be equally relevant – again experience (occasionally complemented by some broad knowledge of local geology) is needed to recognize that some sites are more relevant or "similar". For the construction of a 2D geologic section (or fence diagram), it is not uncommon to connect the same soil type (interpreted by an engineering geologist from disturbed samples) between two adjacent boreholes using a straight line or to insert a question mark with a speculative dashed line to indicate "not sure" if the sequence of soil types does not match between two adjacent boreholes.

There is a strong contrast between judgment-based site characterization that remains relatively unchanged for decades and increasingly sophisticated physics-based modeling of soil-structure interaction such as the material point method. Phoon *et al.* (2022b) opined that: "Currently, numerical analysis is conducted using deterministic inputs and simple soil profiles that are inconsistent with sparse site data. This deterministic mapping (one set of inputs to one set of outputs) does provide valuable physical insights, but cannot support risk-informed decision making on its own without appealing to engineering judgment". Physical insights can be garnered from "what if" parametric studies involving worse credible inputs and/or scenarios, but a formal risk analysis is not possible without uncertainty quantification. Phoon *et al.* (2022c) proposed a "data first practice central" agenda called data-centric geotechnics that seeks to reduce this gap between data and decision making *at a specific site*. While current judgment-based solutions have been effective, they do not benefit from digital technologies. In reference to the increasing amount of information generated by digital technologies, Mitchell and Kopmann (2013) observed that the "evaluation of its validity and importance, and deciding which of it should be included was one of the major challenges faced by the authors, and it provided an excellent example of the 'information overload' problem." To the authors' knowledge, there is no human-in-the-loop decision making mediated by machine learning in geotechnical engineering. The agenda for data-centric geotechnics is underpinned by three core elements: (1) data-centricity, (2) fit for (and transform) practice, and (3) geotechnical context (Phoon *et al.* 2022c).

With the above agenda in mind, Phoon and Zhang (2023) opined that an engineer will be particularly interested in "data quantity and quality (how much data is enough? what type of data is useful?), learning bias (how general is the algorithm?), computational cost (are training and inference costs affordable?), explainability (can the predictions be understood by humans?), and decision making (can judgment be utilized?)". Given the nascent stage of development in data-centric geotechnics, it is not surprising that these questions remain largely unanswered.

In the context of data-driven site characterization (DDSC), Phoon *et al.* (2022c) presented the following seven aspects in Project DeepGeo that are deemed worthy of further study: (1) inputs are restricted to actual real-world site investigation data, (2) data from other sites are routinely used to supplement limited site-specific data in practice based in large part on engineering judgment and appreciation of regional geology, but there is no fully satisfactory machine learning method to do this for BID, (3) human experience cannot be incorporated in any machine learning method thus far to produce continuous improvement in the algorithm, *i.e.,* there is no virtuous machine-human decision cycle (Phoon *et al.* 2022a), (4) machine learning methods remain ''black box", and the lack of ''explainability" of the outputs may not inspire sufficient confidence in an engineer who takes responsibility for the decision, (5) generality of data-driven inferences beyond the scope of the training dataset is unknown, (6) ''bad" data that lead to wrong decisions should exist for a sufficiently large BID, but standard outlier analysis may not be applicable to sparse and incomplete spatially varying datasets, and (7) cost of computing 3D stratigraphic realizations at true scale (*e.g.,* a ''medium" benchmarking problem 40 m long × 20 m wide × 20 m deep cuboid proposed by Phoon *et al.* (2022d) consistent with observed data can be excessive, including the need to re-train from scratch for every new site.

Research in DDSC has gained momentum in recent years (*e.g.,* Huang *et al.* 2018; Cao *et al.* 2019; Ching and Phoon 2019; Wang *et al.* 2019; Ching and Phoon 2020a; Ching and Phoon 2020b; Shuku *et al.* 2020; Ching *et al.* 2021a; Ching *et al.* 2021b; Shi and Wang 2021; Wang *et al.* 2021; Xiao *et al.* 2021a; Xiao *et al.* 2021b; Xu *et al.* 2021; Yoshida *et al.* 2021; Ching *et al.* 2022; Sharma *et al.* 2022; Wu *et al.* 2022; Ching *et al.* 2023a; Shuku and Phoon 2023). It is clear that a number of papers remain method-centric and the contributions are not situated within a broader agenda such as data-centric geotechnics. The gaps between research and practice are less evident in the absence of an agenda that emphasizes "fit for practice". As such, these gaps may not attract the attention they deserve. The directions for future research are likely to be less impactful if transforming practice in a major way is not regarded as a key impetus in the agenda.

The purpose of this paper is to review some research outcomes in the context of data-centric geotechnics using the site recognition challenge as an example. There is no pathway for DDSC to progress without exploiting big data beyond the sparse data available at the target site of interest (data-driven methods need a reasonable amount of data for training), but it is necessary to combine data in a manner sensitive to site differences and even more ideally, sensitive to the accumulated experience and knowledge in geotechnical engineering. Hence, this site recognition challenge is fundamental to DDSC. Recent research progress in this direction is deserving deeper examination particularly in two aspects of central interest to data-centric geotechnics highlighted above by Phoon *et al.* (2022c) and Phoon and Zhang (2023): (1) generality of inferences beyond the scope of the training set and (2) "explainability" of the inferences. These aspects are illustrated using a generic clay property database CLAY/10/7490 (Ching and Phoon 2014) and a target site at Onsøy, Norway (Lacasse and Lunne 1982).

2. SITE RECOGNITION CHALLENGE

The challenge is to quantify "site uniqueness", directly or indirectly, so that sparse site-specific data can be supplemented by big indirect data (BID) to produce a *quasi-site-specific* model. It goes without saying that the statement "two sites are unique" or conversely "two sites are similar" is meaningful only in the probabilistic sense. The value of such a model is that it is less biased than a generic model (containing abundant data from many sites) and less imprecise than a site-specific model (containing sparse data from one site). This idea is not new as geotechnical and rock engineers have been relying on data from similar sites to inform

their understanding of a current site. One simple example is the construction of a correlation to estimate a design parameter from a field test parameter, say to estimate the undrained shear strength from the cone tip resistance. A more generic name is a transformation model (Phoon and Kulhawy 1999a; Phoon and Ching 2017). In practice, site effects are broadly appreciated based on geology, soil mechanics, and experiences at similar sites, rather than characterized quantitatively through a detailed multivariate analysis of the site data. It is routine for engineers to identify the geologic formations in a soil report and make the broad connection that soil properties in the same geologic formation can be used for design even if they are found at different sites. Singapore's geology is divided into the Kallang formation (soft clays), Bukit Timah Granite, Jurong formation (sedimentary), Old Alluvium, *etc.* (http://www.srmeg.org.sg/docs/N13072012_2.pdf). When an engineer encounters the Kallang formation at the current project site, he/she can refer to data from past projects that encountered this formation to supplement limited site-specific data at the current project site. Phoon *et al.* (2004) recommended a cone factor for the estimation of the overconsolidation ratio from the cone tip resistance for the Upper Marine Clay (Kallang formation) in Singapore. For the undrained shear strength versus cone tip resistance transformation model, the cone factor also can be expected to be similar at similar sites. The cone factor can vary between 4.5 and 75 (Djoenaidi 1985), although some differences may be attributable to measurement errors (Kulhawy and Mayne 1990). Ching and Phoon (2012a) conducted an extensive analysis of site effects for piezocone (CPTU) based transformation models using a global CPTU database covering 38 sites worldwide. Although the current geology-based approach can offer qualitative insights on site effects, it is restricted to the localities that an engineer practices in and gains experience from. Big data containing soil records worldwide cannot inform this approach easily.

The typical caveat included in design guides would include a general statement such as "caution must always be exercised when using broad, generalized correlations of index parameters or in-situ test results with soil properties. The source, extent, limitations of each correlation should be examined carefully before use to ensure that extrapolation is not being done beyond the original boundary conditions. 'Local' calibrations, where available, are to be preferred over the broad, generalized correlations" (Kulhawy and Mayne 1990). Notwithstanding this sensible caveat, the engineer is typically left with no recourse but to use these generalized correlations/transformation models in the absence of "local" versions and lack of knowledge of the "source, extent, limitations". A local or site-specific transformation model is commonly unavailable, because there is insufficient data to construct a model of sufficient precision to be useful for decision making (estimate design parameter). Statistical uncertainty is very large when the sample size is small.

The development of a purely data-driven approach that can construct a quasi-site-specific model from site-specific data and similar sites in BID remains an outstanding and very difficult technical challenge, because real site data is "ugly". Phoon *et al.* (2019) presented a useful mnemonic, MUSIC-X (Multivariate, Uncertain and Unique, Sparse, Incomplete, and potentially Corrupted with "X" denoting the spatial/temporal dimension) to highlight seven common "ugly" attributes in real site data. Phoon *et al.* (2022) extended MUSIC-X to MUSIC-3X to cover 3D spatial variability that is more typically exhibited in a real site. "M" and "X" are associated with two distinct types of correlations commonly termed as "cross correlations" and "auto correlations", respectively. Research on constructing a quasi-site-specific transformation model from MUSIC-3X data has been initiated recently only. The metric for distinguishing two sets of MUSIC-3X data is not the same as the classical statistical distance between two probability distributions (Sharma *et al.* 2022). Classical statistical distance such as the Mahalanobis distance, Bhattacharyya distance, or Kullback-Leibler divergence is applicable to ideal data only.

Table 1 summarizes some data-driven methods that attempt to account for site uniqueness in constructing a quasi-site-specific transformation model. The exception is the classical probabilistic multiple regression method (Ching and Phoon 2012b). It regards all site data as belonging to a single generic population (G), *i.e.,* it assumes there is no inter-site variability. It is included in Table 1 for completeness, to serve as a baseline method, and because it is widely applied in practice. The other five methods view site-specific data (S) as distinct from G and attempt to combine S and G based on different models for inter-site variability. Only two methods are "explainable" in the sense that similar records (Ching and Phoon 2020a) or similar sites (Sharma *et al.* 2022) are explicitly identified and hence the engineer can perform a "reality check" on the identified sites and understand how additional data from these sites influenced the quasi-site-specific model and the predictions derived from this model. The majority of the machine learning methods are black boxes – the engineer does not understand the process leading to the predictions. Despite this limitation, they have been shown to reduce bias and imprecision compared to the classical probabilistic multiple regression method within the scope of the training dataset (Ching and Phoon 2019, 2020a; Ching *et al.* 2020, 2021b; Sharma *et al.* 2022). Although the broad contours of the site recognition challenge have been outlined (Phoon *et al.* 2022a), many detailed aspects have not been studied or only studied in a limited way. Let $S = T \cup V$. Assume T is a training set of size n and V is a validation set of size m. There are 2 types of T and V: (1) internal validation: V_1 is fully contained within T_1 and (2) external validation: V_2 is fully outside of T_2. An example of Type 1 is T_1 covering sands with relative density between 20% and 80% (loose to dense) and V_1 covering sands with relative density between 40% and 80% (medium to dense). G is a generic set of size p. S is not part of G and it is typical for $p >> (n + m)$. The following problems are worthy of further study:

1. **"Best" quasi-site-specific model** – Let G_2 be a subset of G that is selected as "similar" to T_1. Assume G_2 can be restricted to size m. Let M_1 be the multivariate PDF constructed from $T_1 \cup V_1$. M_1 is the actual site-specific model. Let M_2 be the multivariate PDF constructed from $T_1 \cup G_2$. M_2 is the quasi-site-specific model. The "best" model can be defined as M_2 close to M_1. A weaker criterion is to compare M_1 and M_2 over one or several pairs of soil parameters and not restricting the size of G_2. T_1 needs to be large enough to train the models and V_1 needs to be larger than T_1 to discern the difference between M_1 and M_2. This proposed criterion can be used to rank the performance of quasi-site-specific models such as those in Table 1.
2. **Extrapolation test (external validation)** – Let G_2 be a subset of G that is selected as "similar" to T_2. The performance of $M_2 = T_2 \cup G_2$ *outside the scope of* T_2 can be validated using V_2. All data-driven methods must be validated internally

Table 1 Methods to construct a quasi-site-specific transformation model from MUSIC-3X data and BID

Method	Big indirect data	Site-specific data	Quasi-site-specific model M	Site uniqueness	Explainable
Probabilistic multiple regression (Ching and Phoon 2012b)	G	G	M = G	No	
Hybrid (Ching and Phoon 2019)	G	S	$M = G \times S$	Yes, but empirical	No
Record similarity (Ching and Phoon 2020a)	$G = (G_1, G_2)$	S	Select G_2 (list of records) ≈ S $M = S + G_2$	Yes, at record level	Yes
2-step Gibbs (Ching *et al.* 2020)	G_2	S	G_2 (regional) ≈ S M = G_2 updated by S	Maybe	No
Hierarchical Bayesian model (Ching *et al.* 2021b)	$G = (G_1, G_2, \ldots, G_n)$	G_{n+1}	M = G updated by G_{n+1}	Yes, at site level	No
Site similarity (Sharma *et al.* 2022)	$G = (G_1, G_2)$	S	Select G_2 (list of sites) ≈ S $M = S + G_2$	Yes, at site level	Yes

using V_1. To the authors' knowledge, there is no theoretical formulation that will assure performance against V_2. There are different constraints to reduce over-fitting in a data-driven method, but the relationship between overfitting and extrapolation (generalizability) is unclear. The authors believe overfitting is mainly intended to improve internal validation, rather than to address external validation explicitly. A reasonable conjecture is that physics-informed data-driven methods may perform better in this extrapolation test (Tao *et al.* 2023). In machine learning, generalizability is studied using meta-learning (learning how to learn). There is no exploration of meta-learning and metadata to generalize algorithms for geotechnical engineering at this point in time.

3. **Quasi-regional clustering** – Ching *et al.* (2020) showed that a municipal database in Shanghai (SH-CLAY/11/4051) is more effective in supporting quasi-site-specific predictions at a target site in Shanghai than a global clay database (CLAY/10/7490). This municipal database can be regarded as a regional subset/cluster in a global database that is defined by a geographical similarity measure (sites located in Shanghai). One can imagine four strategies to make a quasi-site-specific prediction based on the premise of a "regional" advantage:

   (1) Use entire global database with every record assigned an equal weight independent of its similarity to the target site.

   (2) Use entire global database with every record assigned a weight based on its similarity to the target site.

   (3) Use a subset of the global database ("quasi-regional" database) with every record assigned an equal weight

   (4) Use a subset of the global database ("quasi-regional" database) with every record assigned an unequal weight

There are many clustering algorithms that may be useful for identifying the "quasi-regional" cluster nearest to the target site (Xu and Tian 2015). The "quasi-regional" cluster is based on similarity in the database features that may include geology, soil behavior, geographic location, and others. Wu *et al.* (2022) may be the first to study the usefulness of a "quasi regional" cluster for DDSC.

4. **Explainability** – A site is defined as a group of records within a project boundary. M_2 is "explainable" if the records or sites in G_2 can be explicitly retrieved from G and thus can be made accessible to the engineer for "reality check". In more general terms, explainability concerns revealing the inner workings of the "black box" machine learning code that led to the output prediction or any other mechanism that enhance human understanding of the process. As noted above, the engineer is responsible for the final design.

5. **Transferability** – Let H be a second generic population. It can be distinct from G, it can share some common sites with G, or it can be G enlarged by additional sites ($G \subset H$). For a given site S, a new H_2 (new set of similar sites) can be identified by training H using T. However, this is computationally costly. One can expect G to expand with the addition of new sites on a regular basis. Learning is deemed "transferable" if past training outcome (G_2) can inform the determination of H_2, thus avoiding brute force repeated training.

6. **Compatibility** – G is currently restricted to a single soil type such as clay (Ching *et al.* 2016). One example is CLAY/10/7490 (Ching and Phoon 2014). However, S can contain records belonging to different soil types, because of the presence of different soil layers at different depths. All the methods listed in Table 1 require S and G to be compatible, namely only layers belonging to same soil type as G can be analyzed.

This paper explores the second and third aspects using one set of site-specific data and one BID. Needless to say, the above list is not exhaustive, but intended to illustrate the range of largely unexplored questions underlying the site recognition challenge.

3. BIG INDIRECT DATA (BID)

Phoon *et al.* (2019) referred to big data as *indirect* to emphasize the point that big data exists in geotechnical engineering, but it is not directly relevant to one specific project at one specific site. Generic soil property databases (Phoon *et al.* 2016; Ching *et al.* 2016) and load test databases (Phoon and Tang 2019a; Tang and Phoon 2021) are examples of Big Indirect Data (BID). BID is

denoted as G in Table 1. These databases have been made available by ISSMGE TC304 for research: http://140.112.12.21/issmge/tc304.htm?=6. A clay property BID consists of multivariate data from many sites. Table 2 shows a typical example of multivariate data at a single site. It is "Sparse" as it has 9 records, and it is "Incomplete" as it has empty cells shaded in grey. The data from many clay sites can be combined into a BID. One example is CLAY/10/7490 (Ching and Phoon 2014). It contains 7490 records from 251 studies carried out in 30 countries, clearly much larger than Table 2. Each record contains 10 common clay properties:

1. Y_1 = liquid limit (LL)
2. Y_2 = plasticity index (PI)
3. Y_3 = liquidity index (LI)
4. Y_4 = normalized vertical effective stress (σ'_v/P_a)
5. Y_5 = sensitivity (S_t)
6. Y_6 = pore pressure ratio [$B_q = (u_2 - u_0)/(q_t - \sigma_v)$]
7. Y_7 = normalized preconsolidation stress (σ'_p/P_a)
8. Y_8 = undrained strength ratio (s_u/σ'_v)
9. Y_9 = normalized cone tip resistance [$q_{t1} = (q_t - \sigma_v)/\sigma'_v$]
10. Y_{10} = effective cone tip resistance [$q_{tu} = (q_t - u_2)/\sigma'_v$]

in which σ'_v = vertical effective stress; σ'_p = preconsolidation stress; P_a = atmospheric pressure = 101.3 kPa; q_t = (corrected) cone tip resistance; σ_v = vertical total stress; u_2 = porewater pressure directly behind the cone; u_0 = hydrostatic pore pressure; s_u = undrained shear strength. If a record is full, *i.e.,* there are no missing values (or missing test results), this record is complete. A BID is defined as complete only if every record is complete. Such a complete BID is rare in geotechnical engineering. It is not unusual for an entire column to be empty as shown in Table 3 (rock mass properties). The percentage of BID completeness is defined as (number of filled values)/[(number of properties)×(number of rows)]. The percentage completeness for CLAY/10/7490 is 34%. For a BID containing rock mass properties, ROCKMass/9/5876, the percentage completeness is 29% (Ching *et al.* 2021a). A list of soil/rock property BIDs and percentage completeness is given in Table 4. A list of foundation load test BIDs is given in Table 5.

Classical frequentist methods cannot be applied to the MUSIC-X data in Table 2, because it is sparse, incomplete, and there are spatial correlations between records measured at different depths. BID is not sparse, but it remains incomplete (refer to last column of Table 4) and it is not homogeneous due to inter-site variability ("U" or "Unique" attribute). For example, it is tempting to model the data at each site using a multivariate probability density function (PDF) and to quantify site uniqueness using a statistical distance between two site PDFs. Classical statistical distances such as the Kullback–Leibler divergence, Bhattacharyya distance, and Mahalanobis distance are available, but they do not apply to MUSIC-X or MUSIC-3X data as mentioned above. There is no frequentist method to construct a site PDF from MUSIC-X data in the first place. It is rarely appreciated that the conventional approach to calculate a correlation coefficient from a bivariate dataset cannot be extended to an incomplete multivariate dataset. A necessary positive definite property of the correlation matrix cannot be guaranteed (Ching and Phoon 2015). The only practical approach available thus far is Bayesian, because of the additional support from prior distributions (Ching and Phoon 2019).

Table 2 Site investigation data for a site in Onsøy (Norway) (Source: Lacasse and Lunne 1982)

Index	Depth (m)	LL (Y_1)	PI Y_2)	LI (Y_3)	σ'_v/P_a (Y_4)	σ'_p/P_a (Y_5)	s_u/σ'_v (Y_6)	S_t (Y_7)	B_q (Y_8)	q_{t1} (Y_9)	q_{tu} (Y_{10})
1	1.0	56.2	20.0	1.54	0.06	0.85	2.03	6	0.16	29.11	25.57
2	1.9	50.2	18.1	1.82	0.12	0.60	0.91	14	0.24	17.69	14.58
3	3.5	59.9	30.5	0.93	0.22	0.48	0.48	15	0.30	10.52	8.41
4	5.2	56.8	22.9	1.07	0.32	0.45	0.37	7	0.35	7.70	6.11
5	7.6	66.3	31.5	0.87	0.47	0.54	0.24	14	0.47	5.89	4.25
6	9.5	65.1	29.6	0.97	0.58		0.25	12	0.41	6.19	4.74
7	10.8	74.4	36.1	0.81	0.65	0.84	0.25	9	0.46	5.93	4.31
8	13.4	71.4	35.8	0.87	0.81	1.05	0.24		0.47	5.95	4.24
9	16.3	72.7	34.7	0.76	0.99	0.99	0.24		0.55	6.13	3.88

Note: LL = liquid limit; PI = plasticity index; LI = liquidity index; σ'_v = vertical effective stress; σ'_p = preconsolidation stress; P_a = atmospheric pressure = 101.3 kPa; s_u = undrained shear strength; s_u = the in-situ undrained shear strength mobilized in embankment and slope failures (Mesri and Huvaj 2007); S_t = sensitivity q_{t1} = normalized cone tip resistance = $(q_t - \sigma_v)/\sigma'_v$, where q_t = (corrected) cone tip resistance and σ_v = vertical total stress; q_{tu} = effective cone tip resistance = $(q_t - u_2)/\sigma'_v$, where u_2 = porewater pressure directly behind the cone; B_q = pore pressure ratio = $(u_2 - u_0)/(q_t - \sigma_v)$, where u_0 = hydrostatic pore pressure

Table 3 Site data for the İzmir subway site, Turkey (10 out of 32 records shown for illustration) (Kıncal and Koca 2019)

No.	RQD	RMR*	Q	GSI	E_m (GPa)	E_{em} (GPa)	E_{dm} (GPa)	E_i (GPa)	σ_{ci} (MPa)
1	5	25.5			0.11			4.2	26.0
2	18	35			0.238			7.0	39.5
3	32	41.5			0.83			10.8	72.0
4	35	41			0.564			9.0	62.0
5	37.4	41			0.72			10.7	65.2
6	24	36			0.46			10.0	50.0
7	40	39			0.51			8.0	61.0
8	18	35			0.47			8.5	48.5
9	34	35			0.39			7.5	51.2
10	42.5	44.5			1.17			12.3	86.4

* Average of the lower and upper bounds of RMRs reported in Kıncal and Koca (2019)

Note: RQD = rock quality designation; RMR = rock mass rating; Q = Q-system; GSI = geological strength index; E_m = deformation modulus of rock mass; E_{em} = elasticity modulus of rock mass; E_{dm} = dynamic modulus of rock mass; E_i = Young's modulus of intact rock; σ_{ci} = uniaxial compressive strength of intact rock.

4. QUASI-SITE-SPECIFIC TRANSFORMATION MODEL

This section presents some recent research progress in addressing the site recognition challenge. The concept of a "quasi regional" cluster has not been exploited, *i.e.,* the full BID is adopted in existing studies. The objective is to improve the generic transformation model shown in Fig. 1 so that it is less biased and more precise for a specific site. This generic model is widely used in practice (Kulhawy and Mayne 1990). Figures 1(a) and 1b show the data points in CLAY/10/7490 without and with site differentiation, respectively. It is clear from Fig. 1(b) that site effects are not easy to distinguish visually, because the markers for each site do

Table 4 Soil/rock property databases (updated from Ching *et al.* 2021a and Guan *et al.* 2021)

	Database	Reference	Soil/rock parameters	# data points	# sites/studies	% complete
Univariate	CLAY/16	Phoon and Kulhawy (1999a)	γ, γ_d, w_n, PL, LL, PI, LI, ϕ', s_u, s_u^{FV}, q_c, q_t, SPT-N, DMT (A, B), PMT p_L		a	
	SAND/11	Phoon and Kulhawy (1999a)	ϕ', D_r, q_c, SPT-N, DMT (A, B, I_{DMT}, K_{DMT}, E_{DMT}), PMT (p_L, E_{PMT})		b	
	ROCK/8	Prakoso (2002)	γ (or γ_d), n, R, S_h, σ_{bt}, I_s, σ_c, E		c	
	ROCK/13	Aladejare and Wang (2017)	ρ, G_s, I_{d2}, n, w_c, γ, R_L, S_h, σ_{bt}, I_{s50}, σ_c, E, ν		d	
Multivariate	CLAY/5/345	Ching and Phoon (2012b)	LI, s_u, s_u^{re}, σ'_p, σ'_v	345	37 sites	100%
	CLAY/7/6310	Ching and Phoon (2013)	s_u from 7 different test procedures	6310	164 studies	17.7%
	CLAY/6/535	Ching *et al.* (2014)	s_u/σ'_v, OCR, q_{tc}, q_{tu}, $(u_2-u_0)/\sigma'_v$, B_q	535	40 sites	100%
	CLAY/10/7490	Ching and Phoon (2014)	LL, PI, LI, σ'_v/P_a, σ'_p/P_a, s_u/σ'_v, S_t, q_{tc}, q_{tu}, B_q	7490	251 studies	34.1%
	FI-CLAY/7/216	D'Ignazio *et al.* (2016)	s_u^{FV}, σ'_v, σ'_p, w_n, LL, PL, S_t	216	24 sites	100%
	JS-CLAY/5/124	Liu *et al.* (2016)	M_r, q_c, f_s, w_n, γ_d	124	16 sites	100%
	JS-CLAY/7/372	Zou *et al.* (2017)	σ_v, σ'_v, q_{tc}, f_s/σ'_v, B_q, V_{s1}, s_u/σ'_v	372	25 sites	100%
	SAND/7/2794	Ching *et al.* (2017)	D_{50}, C_u, D_r, σ'_v/P_a, ϕ', q_{t1}, $(N_1)_{60}$	2794	176 studies	60.0%
	EMI-ROCK/8/26000+	Kim and Hunt (2017)	σ_c, σ_{bt}, ρ, CAI, PPI, cohesion, direction shear, triaxial confining	26000+	–	–
	FG/5/1000	Kootahi and Moradi (2017)	e, w_n, LL, PI, C_c	1000	170 sites	100%
	ROCK/9/4069	Ching *et al.* (2018)	γ, n, R_L, S_h, σ_{bt}, I_{s50}, V_p, σ_{ci}, E_i	4069	184 studies	34.2%
	FG-KSAT/6/1358	Feng and Vardanega (2019)	e, k, LL, PL, PI, G_s	1358	33 studies	91.4%
	SH-CLAY/11/4051	Zhang *et al.* (2020)	LL, PI, LI, e, K_0, σ'_v/P_a, s_u/σ'_v, S_t, q_c/σ'_v	4051	50 sites	39.5%
	CLAY/8/12225	Ching (2020)	LL, PI, w, e, σ'_v/P_a, C_c, C_{ur}, c_v	12225	427 studies	–
	CLAY/12/3997	Ching (2020)	LL, PI, LI, σ'_v/P_a, σ'_p/P_a, s_u/σ'_v, K_0, E_u/σ'_v, B_q, q_{t1}, $N_{60}/(\sigma'_v/P_a)$	3997	237 studies	–
	SAND/13/4113	Ching (2020)	e, D_r, σ'_v/P_a, σ'_p/P_a, K_0, E_{dn}, q_{c1n}, B_q, $(N_1)_{60}$, K_{DMT}, E_{DMTn}, E_{PMTn}, M_{dn}	4113	172 studies	–
	ROCKMass/9/5876	Ching *et al.* (2021a)	RQD, RMR, Q, GSI, E_m, E_{em}, E_{dm}, E_i, σ_{ci}	5876	225 studies	29.3%
	CLAY-C_c/6/6203	Ching *et al.* (2023a)	LL, PI, w_n, e, C_c, C_{ur}	6203	429 studies	61%
	Global-CPT/3/1196	Ching *et al.* (2023b)	q_t, f_s, u_2	1196	59 sites	100%

Note:

Basic – C_u = coefficient of uniformity; D_{50} = median grain size; ρ = density; G_s = specific gravity; γ = unit weight; γ_d = dry unit weight; D_r = relative density; e = void ratio; n = porosity; w_n (or w_c) = water content; PL = plastic limit; LL = liquid limit; PI = plasticity index; LI = liquidity index; GSI = geological strength index;

Stress – σ_v = total vertical stress; σ'_v = effective vertical stress; σ'_p = preconsolidation stress; OCR = overconsolidation ratio; P_a = atmospheric pressure = 101.3 kPa; K_0 = at-rest lateral earth pressure coefficient

Strength – ϕ' = effective friction angle; s_u = undrained shear strength; s_u^{FV} = field vane s_u; s_u^{re} = remoulded s_u; S_t = sensitivity; σ_{bt} = Brazilian tensile strength; σ_{ci} (or σ_c) = uniaxial compressive strength of intact rock

Deformation – C_c = compression index; C_{ur} = unloading-reloading index; modulus; E_u = undrained modulus of clay; E_d = drained modulus of sand; $E_{dn} = (E_d/P_a)/(\sigma'_v/P_a)^{0.5}$; E_{dm} = dynamic modulus of rock mass; E_{em} = elasticity modulus of rock mass; E_i (or E) = Young's modulus of intact rock; E_m = deformation modulus of rock mass; ν = Poisson ratio; M_r = subgrade resilience modulus; M_d = effective constrained modulus determined by oedometer; M_{dn} = normalized $M_d = (M_d/P_a)/(\sigma'_v/P_a)^{0.5}$

Permeability - k = hydraulic conductivity; c_v = coefficient of consolidation

Dynamic - V_p = P-wave velocity; V_s = S-wave velocity; $V_{s1} = V_s(P_a/\sigma'_v)^{0.25}$

Field test - SPT-N = standard penetration test blow count; N_{60} = corrected SPT-N; $(N_1)_{60} = N_{60}/(\sigma'_v/P_a)^{0.5}$; q_c = cone tip resistance; q_t = corrected cone tip resistance; f_s = sleeve frictional resistance; $q_{tc} = (q_t/P_a)/(\sigma'_v/P_a)^{0.5}$; $q_{t1} = (q_t-\sigma_v)/\sigma'_v$ = normalized cone tip resistance; $q_{tu} = (q_t-u_2)/\sigma'_v$ = effective cone tip resistance; $q_{c1n} = (q_c/P_a)/(\sigma'_v/P_a)^{0.5}$; B_q = pore pressure ratio = $(u_2-u_0)/(q_t-\sigma_v)$; $(u_2-u_0)/\sigma'_v$ = normalized excess pore pressure; u_2 = pore pressure behind cone tip; u_0 = hydrostatic pore pressure; PMT (p_L, E_{PMT}) = pressuremeter limit stress, modulus; E_{PMTn} = normalized $E_{PMT} = (E_{PMT}/P_a)/(\sigma'_v/P_a)^{0.5}$; DMT (A, B, I_{DMT}, K_{DMT}, E_{DMT}) = dilatometer A and B readings, material index, horizontal stress index, modulus; E_{DMTn} = normalized $E_{DMT} = (E_{DMT}/P_a)/(\sigma'_v/P_a)^{0.5}$; CAI = Cerchar abrasivity index; PPI = punch penetration index; Q = Q-system; RMR = rock mass rating; RQD = rock quality designation; R = Schmidt hammer hardness (R_L = L-type Schmidt hammer hardness); S_h = Shore scleroscope hardness; I_{d2} = slake durability index; I_s = point load strength index ($I_{s50} = I_s$ for diameter 50 mm)

a = The no. of data groups varies between 2 and 42 depending on the clay parameter. Statistics are calculated at the data group level. The average no. of data points/data group varies between 16 and 564. Details given in Tables 1-3, Phoon and Kulhawy (1999b).

b = The no. of data groups varies between 5 and 57 depending on the sand parameter. Statistics are calculated at the data group level. The average no. of data points/data group varies between 15 and 123. Details given in Tables 1-3, Phoon and Kulhawy (1999b).

c = The no. of data groups varies between 30 and 174 depending on the rock parameter with no differentiation of rock type [igneous (intrusive, extrusive, pyroclastic), sedimentary (clastic, chemical), metamorphic (foliated, non-foliated)]. Statistics are calculated at the data group level. The average no. of data points/data group varies between 3 and 161 for σ_c (Prakoso, 2017). Details given in Table 4.4, Prakoso (2002).

d = The no. of data groups varies between 2 and 47 depending on the rock parameter and rock type (igneous, sedimentary, or metamorphic). Statistics are calculated at the data group level. The average no. of data points/data group varies between 7 and 92. Details given in Tables 2-4, Aladejare and Wang (2017).

Table 5 Foundation load test databases in Tang and Phoon (2021)

Database/Reference	Limit state	Soil type	# load tests	Pile geometry		Soil parameters
				B (m)	D/B	
NUS/ShalFound/919 (Tang *et al.* 2020)	Bearing	Clay	56	0.3-5	0-5.7	s_u = 9-200 kPa
		Sand	427	0.25-7	0-6.1	ϕ = 26-53°
	Tension	Clay	123	0.31-3.05	0.8-13.2	s_u = 15-300 kPa
		Sand	313	0.1-2.5	0.5-14.5	ϕ = 30-49°
NUS/ShalFound/Punch-Through/31 (Tang and Phoon 2019a)	Punch-through	Sand-over-clay	31	0.8-3	0.5-3	ϕ_{cv} = 32° D_r = 88% s_u = 8.7-85.9 kPa
NUS/Spudcan/Punch-Through/212 (Tang and Phoon 2019a)	Punch-through	Multi-layer clays with sand	140	3-20	0.16-1.17	ϕ_{cv} = 31-34° D_r = 44-99% s_u = 7.2-44.8 kPa
		Multi-layer clays with stiff layer	72	3-12		s_u = 3-50 kPa ρ = 0-2.6 kPa/m
NUS/DrilledShaft/542 (Tang *et al.* 2019)	Bearing	Clay	64	0.32-1.52	1.6-56	s_u = 41-256 kPa
		Sand	44	0.35-2	5.1-59	ϕ = 30-41°
		Gravel	41	0.59-1.5	6.2-30	ϕ = 37-47°
	Tension	Clay	32	0.36-1.8	3.4-55	s_u = 21-250 kPa
		Sand	30	0.3-1.31	2.5-43	ϕ = 30-45°
		Gravel	109	0.43-2.26	1.77-17.3	ϕ = 42-48°
NUS/DrivenPile/1243 H section (Tang and Phoon 2018a; Phoon and Tang 2019b)	Bearing	Clay	47	0.28-0.41	16-95	N_{SPT} = 5-50
		Sand	52	0.28-0.42	22-110	N_{SPT} = 7-40
		Mixed	50	0.28-0.42	17-85	N_{SPT} = 4-29
NUS/DrivenPile/1243 Tube/box section (Tang and Phoon 2019b)	Bearing	Clay	175	0.1-0.81	7.9-200	PI = 11%-160% OCR = 1-43.2 S_t = 1-17
	Tension		64	0.1-0.81	12-110	PI = 12%-110% OCR = 1-43.2 S_t = 1-8.3
NUS/DrivenPile/1243 Tube/box section (Tang and Phoon 2018b)	Bearing	Sand	134	0.14-0.76	13-251	ϕ = 30-42° D_r = 15%-93%
	Tension		28	0.25-0.76	19-84	ϕ = 30-42° D_r = 31%-97%
NUS/RockSocket/721 (Tang and Phoon 2021)	End bearing	Rock	270	0.1-2.5	1-31.3	σ_c = 0.5-99 MPa E_m = 7.82-75113 MPa GSI = 7.5-95 RQD = 20%-100%
NUS/RockSocket/721 (Tang and Phoon 2021)	Shaft shearing	Rock	544	0.2-3.2	0-19.5	σ_c = 0.4-99 MPa E_m = 24-19844 MPa GSI = 50-70 RQD = 0-100%
NUS/HelicalPile/1113 (Tang and Phoon 2018c 2020)	Bearing	Clay	270	0.21-1.02	6-74	$s_u \leq$ 305 kPa
		Sand	181	0.21-1.02	6-110	ϕ = 30-45°
	Tension	Clay	165	0.21-0.91	12-48	$s_u \leq$ 300 kPa
		Sand	121	0.21-0.91	10-62	ϕ = 30-45°

Note: B = foundation diameter; D = foundation embedment depth or thickness of sand layer; s_u = undrained shear strength of clay; ρ = strength gradient; ϕ = friction angle of sand; ϕ_{cv} = constant volume friction angle; D_r = relative density of sand; N_{SPT} = blow count in standard penetration test (SPT); PI = plasticity index; OCR = overconsolidation ratio; S_t = soil sensitivity index; σ_c = uniaxial compressive strength of rock; E_m = elasticity modulus of rock; GSI = geological strength index; and RQD = rock quality designation.

not form distinct non-overlapping clusters. Nonetheless, this does not mean clusters do not exist in higher dimensions.

The site-specific data is shown in Table 2 (Onsøy site, Norway). It is divided in two different ways to illustrate the effect of extrapolation:

1. Internal validation: Training dataset T_1 = row (1, 3, 5, 7, 9) and validation dataset V_1 = row (2, 4, 6, 8).
2. External validation: Training dataset T_2 = row (2, 3, 4, 6, 9) and validation dataset V_2 = row (1, 5, 7, 8).

Internal validation is routinely conducted to evaluate the performance of machine learning methods. There is no theoretical guarantee that machine learning methods will be well behaved under external validation. Nonetheless, it is of practical interest to study the potential degradation of performance using (T_2, V_2).

Using the BID in Fig. 1 and the training dataset T_1 or T_2, a quasi-site-specific model can be developed using different methods to predict the 95% confidence interval (CI) for the undrained strength ratio at the corresponding validation dataset (V_1 or V_2). These 95% CI can be compared with the measured undrained strength ratios, which are available but assumed to be unknown at the model training/calibration stage. For illustration, only the bivariate subset [q_{t1}, (s_u/σ'_v)] is used in this study unless stated otherwise. The methods presented below are not restricted to bivariate datasets. However, they are restricted by a common assumption that all soil properties that are typically non-normal in marginal distributions can be transformed to standard normal random variables and these variables constitute a multivariate normal vector (Ching and Phoon 2015).

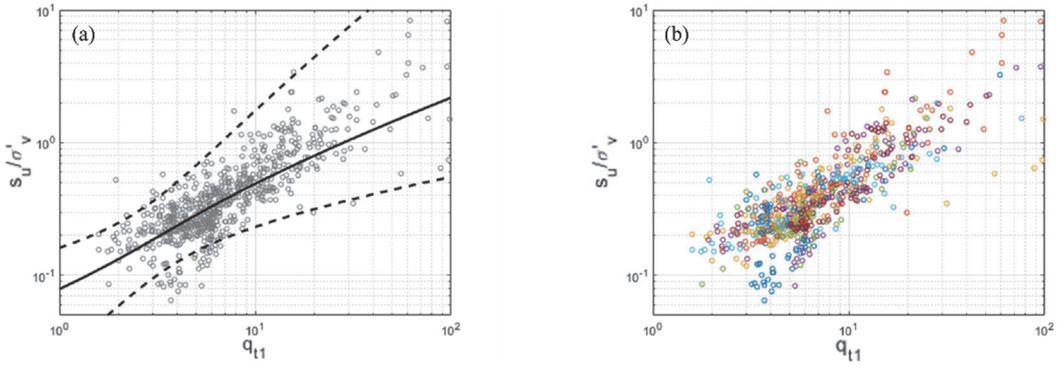

Fig. 1 Generic transformation model for undrained strength ratio (s_u/σ'_v) versus normalized cone tip resistance $q_{t1} = [(q_t - \sigma_v)/\sigma'_v]$ based on CLAY/10/7490 (Ching and Phoon 2014): (a) No site differentiation and (b) Sites differentiated by distinct markers

Five methods are discussed below: (1) probabilistic multiple regression, (2) hybridization, (3) hierarchical Bayesian model, (4) record similarity method, and (5) site similarity method. The first method is generic. The rest are quasi-site-specific. The last two methods are "explainable" in the sense that an engineer can inspect the "similar" sites supporting the construction of the quasi-site-specific (s_u/σ'_v) versus q_{t1} model.

4.1 Probabilistic multiple regression

The probabilistic multiple regression (PMR) was first proposed by Ching and Phoon (2012b) for a complete multivariate soil property database CLAY/5/345. The basic idea is to first convert the physical (Y_8, Y_9) = [q_{t1}, (s_u/σ'_v)] database into a standard normal (X_8, X_9) database and then construct the bivariate normal probability density function (PDF) of (X_8, X_9) based on this standard normal database. For PMR, the training dataset (T_1 or T_2) is simply added to Fig. 1(a) and a new regression model is calculated (Fig. 2(a) or Fig. 2(c)). However, Fig. 2(a) or Fig. 2(c) is almost the same as Fig. 1(a), because the effect of adding 5 records (red markers) to BID [716 records from 72 sites in CLAY/10/7490 contain both q_{t1}, (s_u/σ'_v)] is negligible. Prediction is carried out by reading the values of the 95% CI from Fig. 2(a) or Fig. 2(c) at the four different values of Y_9 in the corresponding validation dataset (Fig. 2(b) or Fig. 2(d)). It can be seen that the actual measured values (yellow markers) fall within the predicted 95% CI. On the average, one expects 1 measured value out of 20 values to fall outside a 95% CI. PMR is the standard approach widely adopted in practice, typically in a bivariate form (Kulhawy and Mayne 1990). It produces a generic model, not a quasi-site-specific model. There is no difference between internal and external validation in this example, because V_1 and V_2 are within the broad coverage of the generic transformation model. The authors are not aware of any method that can modify PMR for site effects, short of an engineer manually selecting "similar" sites before applying PMR. Figure 2 presents a baseline for comparison with the quasi-site-specific models discussed in the next four sections.

4.2 Hybridization

The hybridization method (HYB) was proposed by Ching and Phoon (2019). The basic idea is to first construct the generic PDF model in a standard normal (X_8, X_9) space as done in PMR. A site-specific PDF model is next constructed based on the site-specific training dataset (T_1 or T_2) using Bayesian machine learning (also in standard normal space). As noted above, there is no classical frequentist approach that can do this for the type of data shown in Table 2 or Table 3. The "trained" quasi-site-specific model is a hybrid PDF that is proportional to the product of the generic PDF and site-specific PDF. There is no theoretical basis for this "product" hybridization step. The hybrid PDFs so produced by T_1 and T_2 are shown in Figs. 3(a) and 3(c), respectively. Note that additional values covering a much wider range of $1 < q_{t1} < 100$ are added to T_1 and T_2 to extend the quasi-site-specific model beyond the range of $5.89 < q_{t1} < 29.11$ in Table 2. However, the corresponding values of (s_u/σ'_v) are missing, since they are not measured. Figures 3(a) and 3(c) look similar, because T_1 and T_2 are small datasets containing 5 records, resulting in relatively diffused PDFs with large standard deviations. The product of these diffused PDFs with a generic PDF is likely to be dominated by the generic PDF. Hence, it is not surprising that Figs. 3(a) and 3(c) are also similar to Fig. 1(a) (generic PDF).

For internal validation, hybridization works better than PMR, because the 95% CIs in Fig. 3(b) are smaller than those in Fig. 2(b). For external validation, hybridization seems to produce results no worse than Fig. 2(d). The hybridization method is designed so that the quasi-site-specific model converges to the generic model when the site-specific data becomes uninformative (very sparse and/or very incomplete). As such, the performance of the hybridization method cannot be poorer than that of the generic model even under external validation, unless V_2 lies outside the generic database (CLAY/10/7490). The dashed boxes in Fig. 3 maps the range of [q_{t1}, (s_u/σ'_v)] in the training dataset to the validation figures to highlight the degree of extrapolation.

Engineers frequently asked for guidance on the minimum number of data points needed to construct a reliable local or site-specific transformation model. This is an important practical question. There is an assumption underlying this question that data is univariate and there are no other data attributes affecting reliability. As shown in Tables 2 and 3, actual site data is commonly multivariate in nature. The more complete attributes for multivariate data are summarized as MUSIC-3X. Sparsity or a limited number of data points is only one attribute, and it is not the only attribute that can affect the reliability of a transformation model. Hence, for the general case of MUSIC-3X, there is no simple or general answer, because reliability depends on the transformation model, the number of soil parameters involved, the attributes of the site-specific data (*e.g.,* sparsity, incompleteness, strength of cross/spatial correlations, percentage of corrupted data, *etc.*), and the desired precision (*e.g.,* tolerable size of the 95% confidence interval for a site-specific prediction).

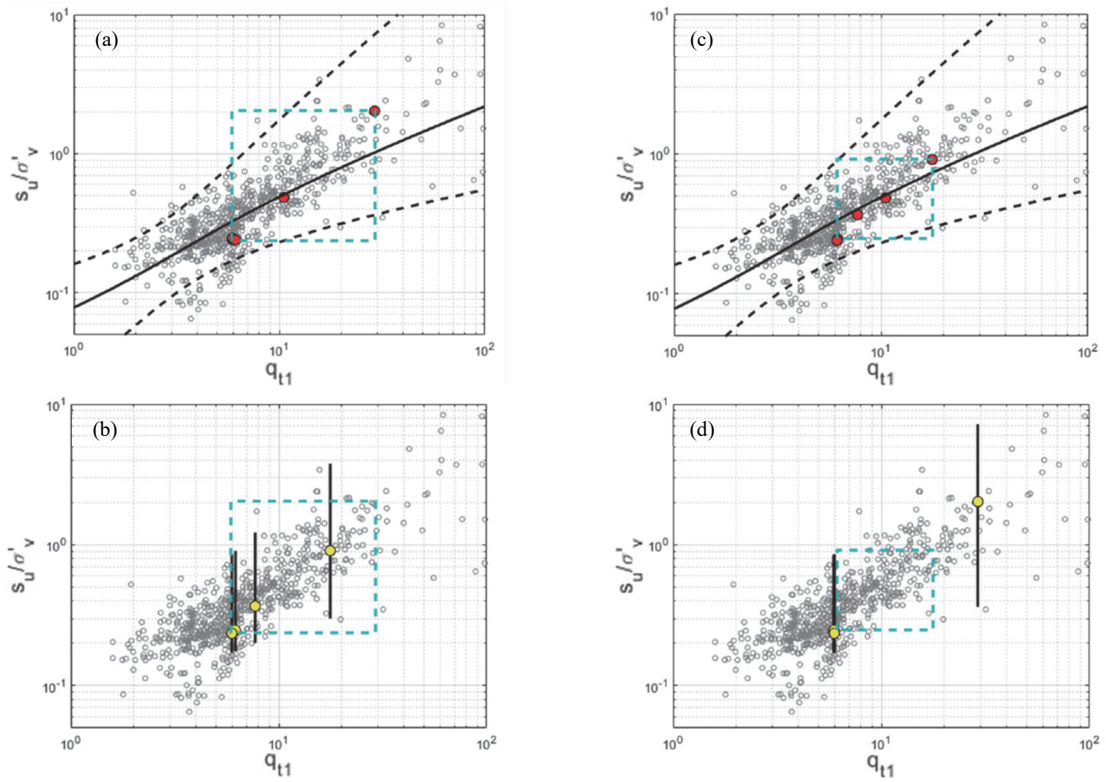

Fig. 2. ***Generic*** **transformation model based on** ***probabilistic multiple regression*** **(PMR): (a) "training" using T_1 (direct combination of T_1 and generic data); (b) internal validation V_1; (c) "training" using T_2 (direct combination of T_2 and generic data); (d) external validation V_2. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

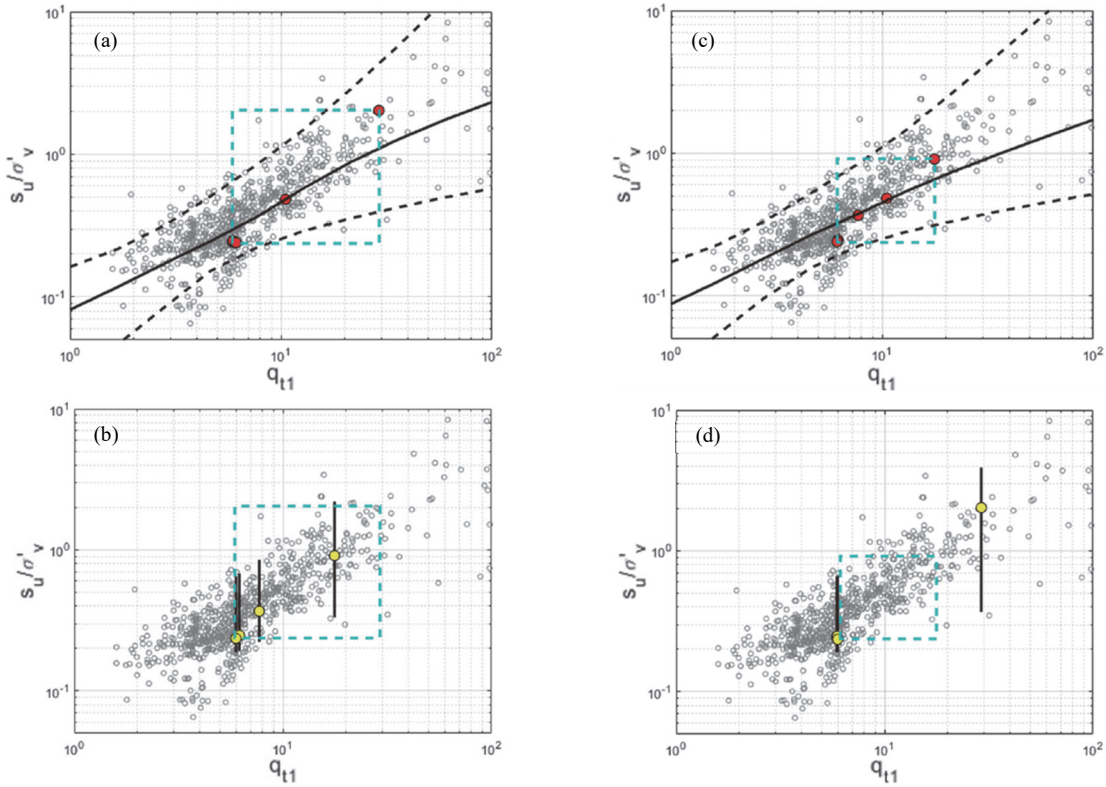

Fig. 3 ***Quasi-site-specific*** **model based on the** ***hybridization*** **method (HYB): (a) "training" using T_1 (product of T_1 and generic PDF); (b) internal validation V_1; (c) "training" using T_2 (product of T_2 and generic PDF); (d) external validation V_2. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

However, it is possible to answer a weaker version of this question: is the entire table of data records (not just one column of numbers) collected in a site investigation program *not* good enough? As a concrete example, is Table 2 not good enough? A possible approach is to apply the hybrid method as described above. Based on Fig. 3(a), it can be seen that the hybrid solution is almost the same as the generic solution shown in Fig. 2(a). Even if the Onsøy clay happens to behave in the same way as a "generic" clay (rather unlikely), this does not affect the conclusion that Table 2 did not provide information beyond what is available in CLAY/10/7490. As far as this specific transformation model is concerned, Table 2 has very limited value to decision making (estimation of s_u/σ'_v) in terms of reducing bias and/or increasing precision. As such, Table 2 is not good enough to construct a local model. It will be useful to provide more guidance to the engineer if he/she wishes to conduct a second site investigation program based on what is known in Table 2, such as: (1) which sub-group of soil parameters should be tested more extensively, (2) how many additional tests are needed for each soil parameter, (3) where are the locations for the additional tests, (4) which new soil parameters should be measured, and (5) what is most cost effective strategy based on the above questions. To the authors' knowledge, there is no data-driven methodology to do this at present, although inverse analysis may provide a partial ingress to this challenge (Phoon *et al.* 2022b).

4.3 Hierarchical Bayesian model

The records for each site can be grouped by a distinct marker as shown in Fig. 1(b). The scatter within one group is called intra-site variability. The scatter between groups is called inter-site variability. If CLAY/10/7490 is structured in this way, a natural approach to construct the quasi-site-specific transformation is to apply the hierarchical Bayesian model (HBM) (Gelman and Hill 2006). However, there are two practical problems in applying HBM to site data. First, it can be computationally tedious. Second, HBM has not been applied to MUSIC-X or MUSIC-3X data. Ching *et al.* (2021b) proposed an analytical HBM for MUSIC data (assuming no spatial variability) based on conjugate priors that is computationally efficient. In comparison to hybridization, HBM has a sound theoretical basis. Although the underlying data structure in CLAY/10/7490 may not follow HBM, particularly an analytical one based on conjugate priors, it is arguably a reasonable first step to study site uniqueness under an established theoretical framework. The quasi-site-specific transformation models based T_1 and T_2 are shown in Figs. 4(a) and 4(c), respectively. Note that Figs. 4(a) and 4(c) are no longer the same. In this sense, HBM is more sensitive to the training dataset. For internal validation, HBM works better than HYB, because the 95% CIs in Fig. 4(b) are smaller than those in Fig. 3(b). For external validation, HBM also seems to work better than HYB with smaller 95% CIs that capture the measured values closer to the medians. More importantly, HBM produces a model that looks different from HYB. It has been pointed out in Section 4.2 that T_1 is not good enough to produce a local (site-specific) model, because HYB looks the same as the generic model in Fig. 2(a). However, T_1 appears to be good enough to change the generic model when it can draw upon the HBM inter-site variability information. It is noteworthy that this is possible, although T_1 only contains 5 records. HYB does not contain inter-site variability information.

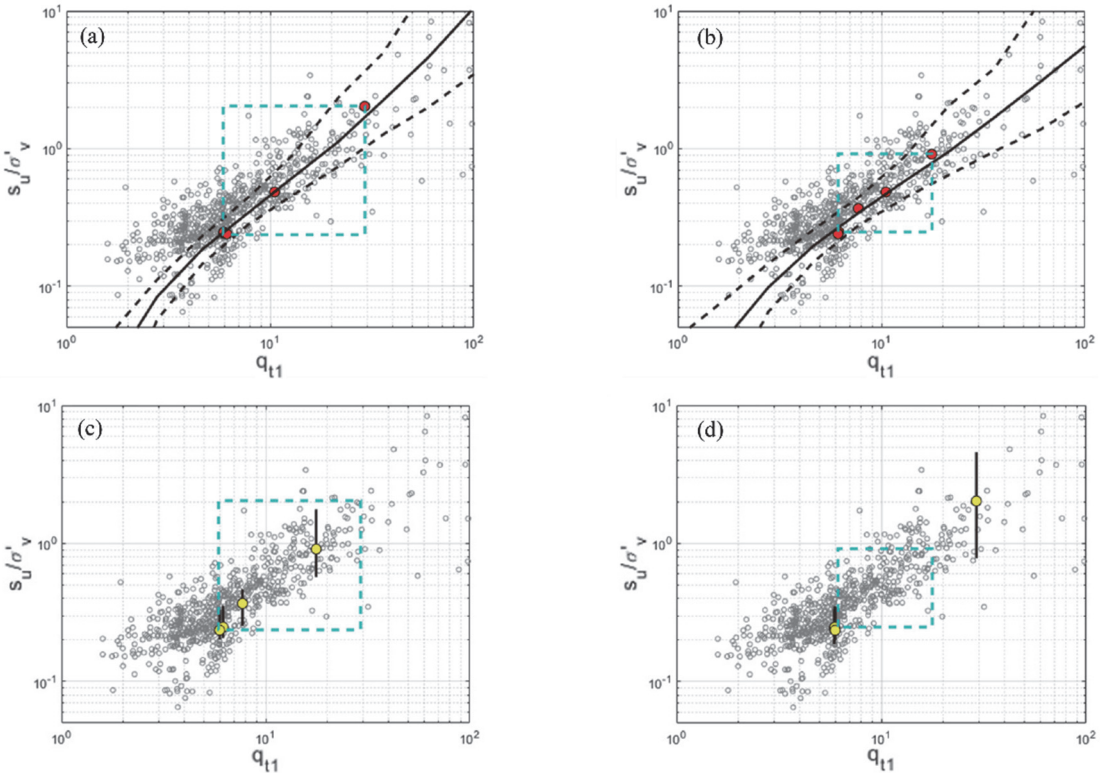

Fig. 4 ***Quasi-site-specific*** **model based on the** ***hierarchical Bayesian model*** **(HBM): (a) training using T_1 (compute T_1 PDF based on HBM hyper-parameters calibrated from CLAY/10/7490); (b) internal validation V_1; (c) training using T_2 (compute T_2 PDF based on HBM hyper-parameters calibrated from CLAY/10/7490); (d) external validation V_2. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

4.4 Record similarity method

Phoon and Zhang (2023) opined that decision making is the responsibility of an engineer (not the operation) and this is ideally carried out if the engineer can understand the inference produced by the operation ("explainable" inference). The key limitation of hybridization and HBM is that they are not "explainable" in the sense that "similar" sites are not identified and the engineer is thus deprived of an opportunity to inspect the list of "similar" sites against his/her experience and knowledge of regional geology ("reality check"). An engineer cannot engage meaningfully in the decision loop without understanding how an inference is arrived at. This is the well-known "black box" problem in machine learning.

Ching and Phoon (2020a) proposed the first approach to address this "explainable site recognition" challenge. The basic idea is to construct a site-specific PDF using Bayesian machine learning as presented in Ching and Phoon (2019). The degree of "similarity" of a *record* in CLAY/10/7490 to the Onsøy training dataset (T_1 or T_2) is proportional to the value of the PDF corresponding to the record. The constant of proportionality changes according to the incompleteness of the record. Based on this adjustment, the similarity of two records with different missing variables (different incompleteness structure) can be compared. A record with "similarity" > 1 means it is more "similar" to the Onsøy training dataset compared to an average record in CLAY/10/7490. These records are highlighted by a triangular marker in Fig. 5.

In contrast to HBM, "similar" records are made known to the engineer for inspection. For example, an engineer can compare the Onsøy data (training and validation) with records with "similarity" > 1 in the Casagrande plasticity chart (Fig. 6(a)). The "similar" records do not cluster around the Onsøy data in the Casagrande plasticity chart. This does not imply that the record similarity method is incorrect, because "similarity" in this example is measured in a bivariate space $(Y_6, Y_9) = (s_u/\sigma'_v, q_{t1})$. The Casagrande plasticity chart measures "similarity" in (LL, PI) or (Y_1, Y_2). Using the classical face recognition problem as an analogue, this means similarity between eyes and nose does not imply similarity between ears and mouth. The results in Fig. 6(a) seem to hint that similarity should be measured by more than two properties. However, it is also known that all sites become distinct if too many properties are included. Figure 6(b) shows the results of measuring "similarity" in a 5-dimensional space $(Y_3, Y_5, Y_6, Y_8, Y_9) = (\mathrm{LI}, \sigma'_p/P_a, s_u/\sigma'_v, B_q, q_{t1})$. There are less "similar" records as to be expected but the degree of clustering around the Onsøy data does not improve. The research question on whether an optimum number and/or optimum list of properties exist for explainable site recognition analysis has not been answered.

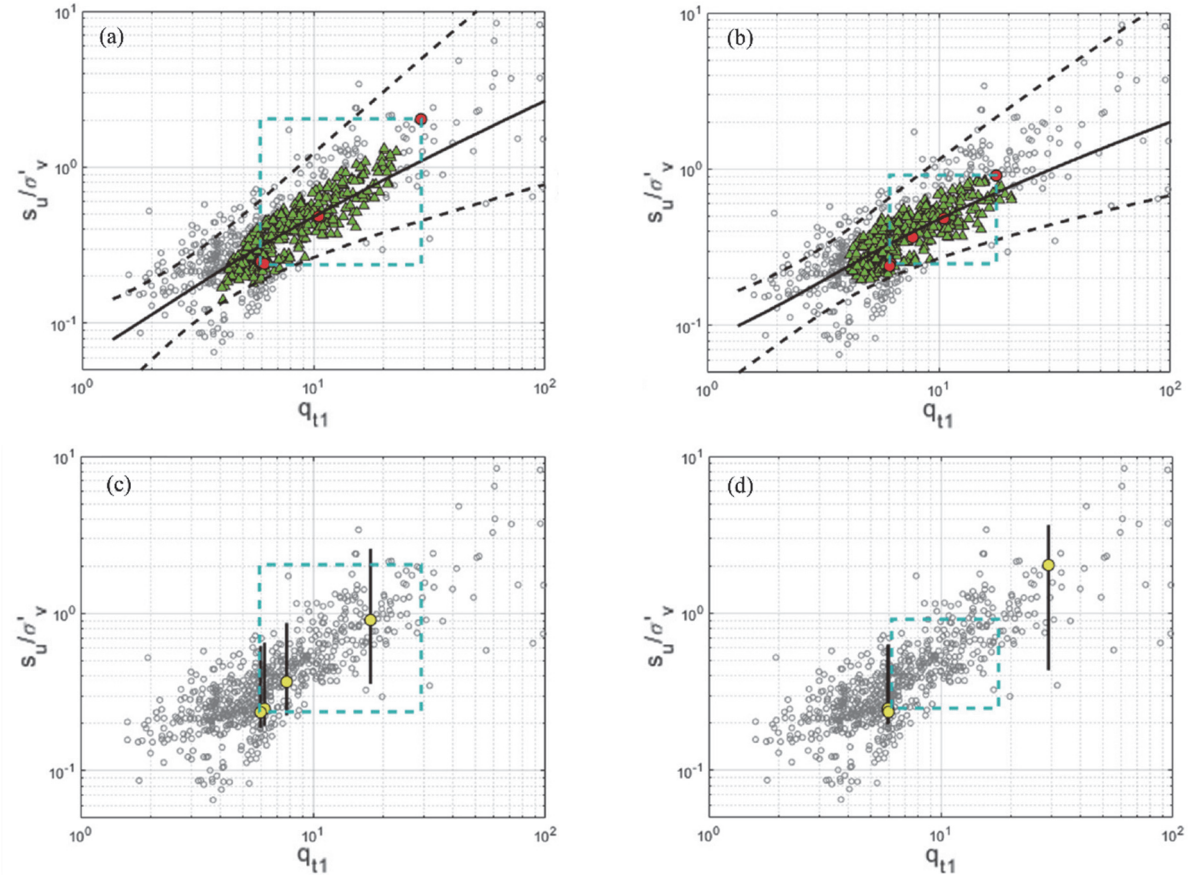

Fig. 5 ***Quasi-site-specific*** **model based on the** ***record similarity*** **method: (a) training using T_1 (weighted regression with each record in T_1 carrying a weight of 1 and each record in CLAY/10/7490 carrying a weight proportional to its record "similarity" to T_1); (b) internal validation V_1; (c) training using T_2 (weighted regression with each record in T_2 carrying a weight of 1 and each record in CLAY/10/7490 carrying a weight proportional to its record "similarity" to T_2); (d) external validation V_2. Records with similarity index > 1 are shown in green triangular markers. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

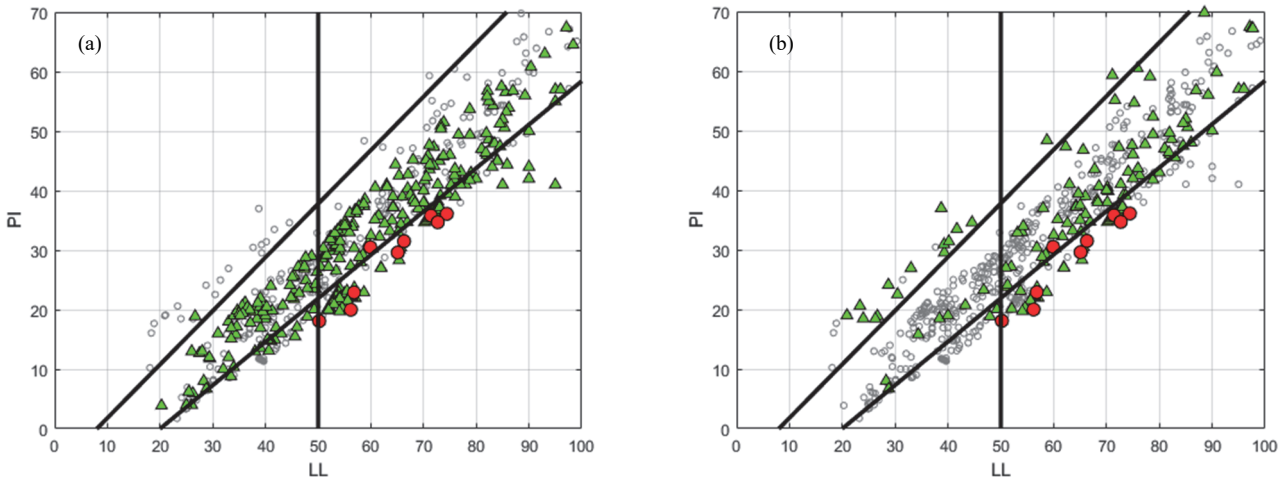

Fig. 6 Comparison between Onsøy site records (red circular marker) and "similar" CLAY/10/7490 records (green triangular marker) using the Casagrande plasticity chart. "Similar" records are defined by a *record similarity* index > 1 computed in: (a) bivariate space (Y_6, Y_9) = (s_u/σ'_v, q_{t1}) and (b) 5-dimensional space (Y_3, Y_5, Y_6, Y_8, Y_9) = (LI, σ'_p/P_a, s_u/σ'_v, B_q, q_{t1})

Finally, a quasi-site-specific transformation model is constructed by performing a weighted regression with T_1 and T_2 and all records in CLAY/10/7490 as shown in Figs. 5(a) and 5(c), respectively. The weight for each record in the Onsøy training dataset (T_1 or T_2) is equal to 1 by construction. The approach to do this is to divide the training set, say T_1, into 2 parts. The first part T1-A contains 4 records. The second part T1-B contains 1 record. T1-A forms the basis to construct a site-specific PDF using Bayesian machine learning. T1-B is considered as an "external" record in a generic database such as CLAY/10/7490, although it is part of the Onsøy site data in actuality. The "similarity" of the single record in T1-B as measured by the site-specific PDF based on T1-A can be computed. It is possible for each record in T1 to be classified as T1-B. Hence, the above exercise can be repeated 5 times and the average "similarity" S_{ave} can be computed. S_{ave} can be viewed as a "self-similarity". The weight for a record in CLAY/10/7490 is defined as its "similarity" divided by S_{ave}. Using this "leave-one-out" approach, the weight of the site-specific data is 1 (approximately) by construction. The weight for a record in CLAY/10/7490 is close to 1 if it is similar to the Onsøy training dataset and close to 0 if it is dissimilar. In this way, it is not necessary to prescribe an arbitrary threshold to divide between "similar" and "dissimilar" sites. The difference between Figs. 5(a) and 2(a) is the use of weights in regression. Figure 2(a) is based on regression with a weight of 1 for all records regardless of similarity.

For internal and external validation, the record similarity method seems to work marginally better than HYB but less well when compared to HBM.

4.5 Site similarity method

The record similarity method (Ching and Phoon 2020) can only compare one record in CLAY/10/7490 with the records in the Onsøy site. In other words, it can consider site grouping at the target site (Onsøy) but it cannot consider site grouping in CLAY/10/7490. Sharma *et al.* (2022) proposed a site similarity method based on HBM to address this limitation. The site similarity measure is essentially the joint density of the Onsøy data based on the PDF of one comparison site in CLAY/10/7490. The PDF is evaluated approximately using HBM. This measure is normalized so that it is unity when the Onsøy data is compared with its own PDF (self-similarity). This site similarity measure elegantly reduces to the classical Kullback–Leibler divergence for complete multivariate data.

Figures 7(a) and 7(c) show the quasi-site-specific models and sites with similarity > 0.1 (triangular markers) corresponding to T_1 and T_2, respectively. The quasi-site-specific models are obtained by weighted regression with weights given by the site similarity method. For internal and external validation, the performance of the site similarity method and record similarity method appears to be comparable. It is more interesting to observe that the site similarity method compares unfavourably to HBM. The explicit identification of "similar" sites has apparently reduced the efficacy of HBM. More research is needed to understand the reasons for this apparent trade-off between explainability and inference (bias and precision).

Figure 8(a) shows that "similar" records identified by the site similarity method do not cluster around the Onsøy data in the Casagrande plasticity chart as well. When site similarity is measured in a 5-dimensional space, the number of "similar" sites decreases significantly (more so than what is shown in Fig. 6(b). A comparison between Fig. 8a and 9a shows that the quasi-site-specific model does change when site similarity is measured in a different space. The application of site similarity in a 5-dimensional space appears to improve inference precision in this Onsøy example (Fig. 9(b)).

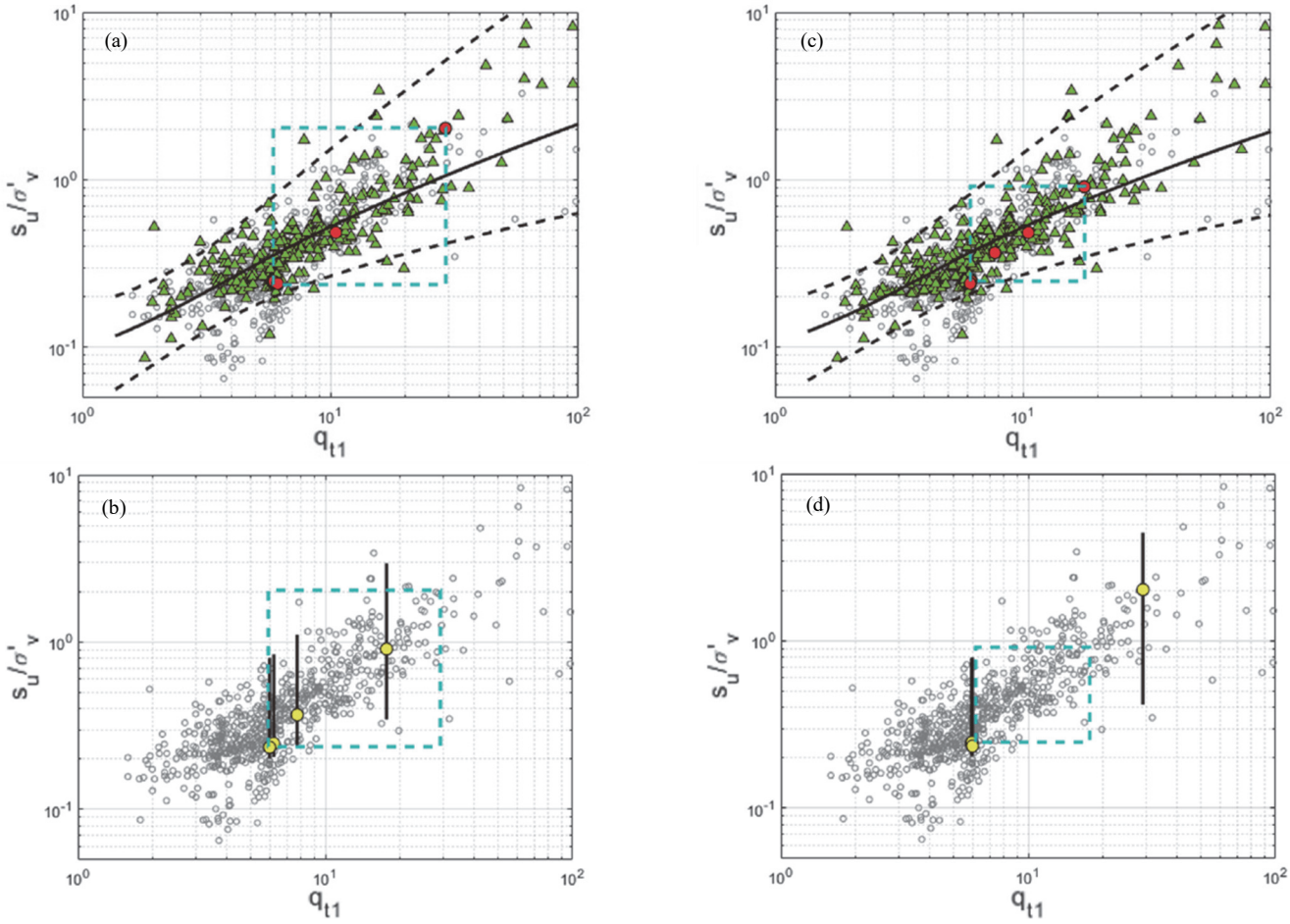

Fig. 7 **Quasi-site-specific model based on the site similarity method: (a) training using T_1 (weighted regression with T_1 carrying a weight of 1 and each site in CLAY/10/7490 carrying a weight proportional to its site "similarity" to T_1); (b) internal validation V_1; (c) training using T_2 (weighted regression with T_2 carrying a weight of 1 and each site in CLAY/10/7490 carrying a weight proportional to its site "similarity" to T_2); (d) external validation V_2. Records with site similarity index > 0.1 are shown in green triangular markers. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

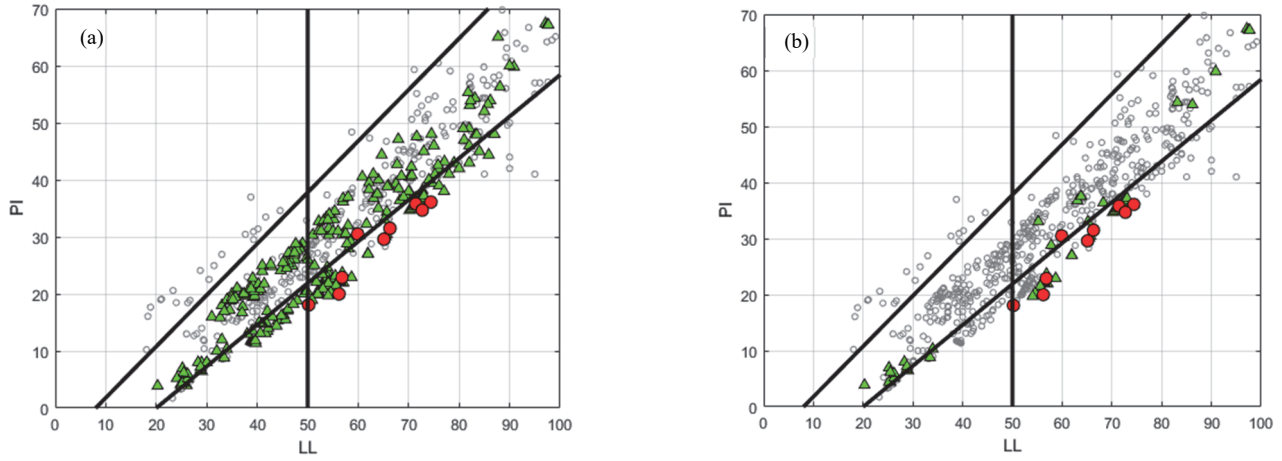

Fig. 8 **Comparison between Onsøy site records (red circular marker) and "similar" CLAY/10/7490 records (green triangular marker) using the Casagrande plasticity chart. "Similar" records are defined by a *site similarity* index > 0.1 computed in: (a) bivariate space (Y_6, Y_9) = (s_u/σ'_v, q_{t1}) and (b) 5-dimensional space (Y_3, Y_5, Y_6, Y_8, Y_9) = (LI, σ'_p/P_a, s_u/σ'_v, B_q, q_{t1})**

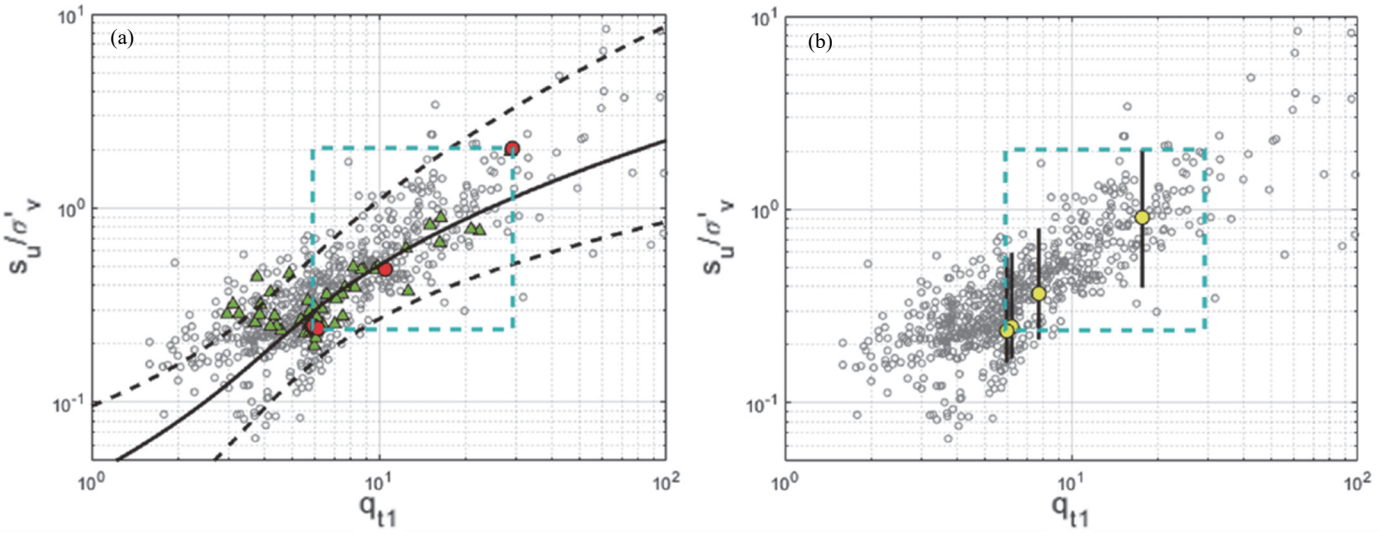

Fig. 9 ***Quasi-site-specific*** **model based on the** ***site similarity*** **method applied in a** ***5-dimensional space*** **(LI, σ'_p/P_a, s_u/σ'_v, B_q, q_{t1}): (a) training using T_1 (weighted regression with T_1 carrying a weight of 1 and each site in CLAY/10/7490 carrying a weight proportional to its site "similarity" to T_1), (b) internal validation V_1. Records with site similarity index > 0.1 are shown in green triangular markers. Onsøy training datasets are shown in red circular marker. Onsøy validation datasets are shown in yellow circular marker**

5. CONCLUSIONS

In the absence of site-specific data, an engineer is compelled to rely on a generic transformation model to estimate a design parameter, such as the correlation between the undrained shear strength and the normalized cone tip resistance. There is rarely sufficient data to establish a site-specific or local correlation. The engineer understands that a generic transformation model is biased when applied to a specific site. He/she relies on engineering judgment to correct this bias approximately. However, there is no judgment possible to correct the generic transformation uncertainty which is too large for a specific site.

Alternately, an engineer may combine site data with data from other similar sites to construct a quasi-site-specific model based on experience and knowledge of geology. No data-driven method is available to construct such a model. This is called the "site recognition" challenge in the data-driven site characterization (DDSC) research agenda. It is a difficult challenge primarily because the attributes of actual site data are MUSIC-3X.

This paper presents four data-driven methods to construct a quasi-site-specific model to estimate the undrained shear strength from the normalized cone tip resistance: (1) hybridization, (2) hierarchical Bayesian model, (3) record similarity method, and (4) site similarity method. The last two methods are "explainable" in the sense that the list of "similar" records or sites supporting the quasi-site-specific model is made known to the engineer. In this way, the engineer can delete records/sites deemed unreasonable, thus fostering a meaningful engagement in the decision making (estimation of the undrained shear strength). The effect of extrapolating the quasi-site-specific model beyond the range of the training dataset is also studied by comparing the performance of these models under routine validation (validation dataset is contained within the training dataset) and under external validation (validation dataset lies outside the training dataset). The hierarchical Bayesian model appears to be the best performing method thus far, but it suffers from a lack of "explainability". The hybridization model is however robust against extrapolation, because it cannot perform worse than the generic model by construction.

More research is needed to: (1) clarify the relationship between data-driven "similarity" and physics-based soil classification (*e.g.,* Casagrande plasticity chart, Robertson CPT-based soil behavior type classification system), (2) ascertain the number and/or type of soil properties needed to produce "similar" sites that will appear as clusters in soil classification charts (if a relationship exists) and/or produce the most accurate quasi-site-specific model, (3) study the robustness of making inferences beyond the range of the training dataset or , and (4) understand the trade-off between explainability and inference accuracy (bias and precision). Six research problems are posed in this paper to illustrate the range of largely unexplored questions underlying the site recognition challenge: (1) "best" quasi-site-specific model, (2) extrapolation test, (3) quasi-regional clustering, (4) explainability, (5) transferability, and (6) compatibility.

ACKNOWLEDGMENTS

This paper is expanded from a keynote paper entitled "The 'site recognition challenge' in data-driven site characterization" presented at the 5th International Conference on New Developments in Soil Mechanics and Geotechnical Engineering, Nicosia, North Cyprus, 30 June to 1 July 2022. The presented results based on a new target site (Table 2) are however original to this paper. The authors thanked Professor Feyza Çinicioğlu and Professor Cavit Atalar, Organising Committee Chairs, for their kind invitation.

The authors would like to thank the members of the TC304 Committee on Engineering Practice of Risk Assessment & Management of the International Society of Soil Mechanics and Geotechnical Engineering for developing the database 304dB used in this study and making it available for scientific inquiry.

FUNDING

The authors received no funding for this work.

DATA AVAILABILITY

This study does not generate new data and/or new computer codes.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Aladejare, A.E. and Wang, Y. (2017). "Evaluation of rock property variability." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **11**(1), 22-41. https://doi.org/10.1080/17499518.2016.1207784

Bond A. and Harris A. (2008). *Decoding Eurocode 7*, Taylor and Francis, London.

Cao, Z., Zheng, S., Li, D.Q., and Phoon, K.K. (2019). "Bayesian identification of soil stratigraphy based on soil behaviour type index." *Canadian Geotechnical Journal*, **56**(4), 570-586. https://doi.org/10.1139/cgj-2017-0714

CEN (2004). *EN 1997-1: Geotechnical design – Part 1: General rules*. Comité Européen de Normalisation.

Ching, J. (2020). *Unpublished Databases*.

Ching, J. and Phoon, K.K. (2012a). "Establishment of generic transformations for geotechnical design parameters." *Structural Safety*, **35**, 52-62. https://doi.org/10.1016/j.strusafe.2011.12.003

Ching, J. and Phoon, K.K. (2012b). "Modeling parameters of structured clays as a multivariate normal distribution." *Canadian Geotechnical Journal*, **49**(5), 522-545. https://doi.org/10.1139/t2012-015

Ching, J. and Phoon, K.K. (2013). "Multivariate distribution for undrained shear strengths under various test procedures." *Canadian Geotechnical Journal*, **50**(9), 907-923. https://doi.org/10.1139/cgj-2013-0002

Ching, J. and Phoon, K.K. (2014). "Transformations and correlations among some clay parameters – the global database." *Canadian Geotechnical Journal*, **51**(6), 663-685. https://doi.org/10.1139/cgj-2013-0262

Ching, J. and Phoon, K.K. (2015). "Constructing multivariate distributions for soil parameters." *Risk and Reliability in Geotechnical Engineering*, CRC Press, 3-76.

Ching, J. and Phoon, K.K. (2019). "Constructing site-specific multivariate probabilistic distribution model by Bayesian machine learning." *Journal of Engineering Mechanics*, ASCE, **145**(1), 04018126. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001537

Ching, J. and Phoon, K.K. (2020a). "Measuring similarity between site-specific data and records from other sites." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, **6**(2), 04020011. https://doi.org/10.1061/AJRUA6.0001046

Ching, J. and Phoon, K.K. (2020b). "Constructing a site-specific multivariate probability distribution using sparse, incomplete, and spatially variable (MUSIC-X) data." *Journal of Engineering Mechanics*, ASCE, **146**(7), 04020061. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001779

Ching, J., Phoon, K.K., and Chen, C.H. (2014). "Modeling CPTU parameters of clays as a multivariate normal distribution." *Canadian Geotechnical Journal*, **51**(1), 77-91. https://doi.org/10.1139/cgj-2012-0259

Ching, J., Li, D.Q., and Phoon, K.K. (2016). "Statistical characterization of multivariate geotechnical data." Chapter 4, *Reliability of Geotechnical Structures in ISO2394*, CRC Press/Balkema, 89-126.

Ching, J., Lin, G.H., Chen, J.R., and Phoon, K.K. (2017). "Transformation models for effective friction angle and relative density calibrated based on a multivariate database of coarse-grained soils." *Canadian Geotechnical Journal*, **54**(4), 481-501. https://doi.org/10.1139/cgj-2016-0318

Ching, J., Li, K.H., Phoon, K.K., and Weng, M.C. (2018). "Generic transformation models for some intact rock properties." *Canadian Geotechnical Journal*, **55**(12), 1702-1741. https://doi.org/10.1139/cgj-2017-0537

Ching, J., Phoon, K.K., Khan, Z., Zhang, D.M., and Huang, H.W. (2020). "Role of municipal database in constructing site-specific multivariate probability distribution." *Computers and Geotechnics*, **124**, 103623. https://doi.org/10.1016/j.compgeo.2020.103623

Ching, J.Y., Phoon, K.K., Ho, Y.H., and Weng, M.C. (2021a). "Quasi-site-specific prediction for deformation modulus of rock mass." *Canadian Geotechnical Journal*, **58**(7), 936-951. https://doi.org/10.1139/cgj-2020-0168

Ching, J., Wu, S., and Phoon, K.K. (2021b). "Constructing quasi-site-specific multivariate probability distribution using hierarchical Bayesian model." *Journal of Engineering Mechanics*, ASCE, **147**(10), 04021069. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001964

Ching, J., Phoon, K.K., Yang, Z.Y., and Stuedlein, A.W. (2022). "Quasi-site-specific multivariate probability distribution model for sparse, incomplete, and three-dimensional spatially varying soil data." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **16**(1), 53-76. https://doi.org/10.1080/17499518.2021.1971256

Ching, J., Yoshida, I., and Phoon, K.K. (2023a). "Comparison of trend models for geotechnical spatial variability: Sparse Bayesian learning vs. Gaussian process regression." *Gondwana Research*. https://doi.org/10.1016/j.gr.2022.07.011

Ching, J., Phoon, K.K., and Wu, C.T. (2023b). "Data-centric quasi-site-specific prediction for compressibility of clays." *Canadian Geotechnical Journal*. https://doi.org/10.1139/cgj-2021-0658

Ching, J., Uzielli, M., Phoon, K.K., and Xu, X.J. (2023c). "Characterizing spatially variable cone tip resistance soundings from a global CPT database." *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, in review.

D'Ignazio, M., Phoon, K.K., Tan, S.A., and Lansivaara, T. (2016). "Correlations for undrained shear strength of Finnish soft clays." *Canadian Geotechnical Journal*, **53**(10), 1628-1645. https://doi.org/10.1139/cgj-2016-0037

Djoenaidi, W.J. (1985). "A compendium of soil properties and correlations." *Thesis for Master of Engineering Science*, University of Sydney, Australia.

Feng, S. and Vardanega, P.J. (2019). "A database of saturated hydraulic conductivity of fine-grained soils: probability density functions." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **13**(4), 255-261.

https://doi.org/10.1080/17499518.2019.1652919

Gelman, A. and Hill, J. (2006). *Data Analysis using Regression and Multilevel/hierarchical Models*. Cambridge University Press.

Guan, Z., Chang, Y.C., Wang, Y., Aladejare, A., Zhang, D.M., and Ching, J. (2021). "Site-specific statistics for geotechnical properties." *TC304 State-of-the-art Review of Inherent Variability and Uncertainty in Geotechnical Properties and Models*, 1-83. https://doi.org/10.53243/R0001

Huang, J., Zheng, D., Li, D., Kelly, R., and Sloan, S. (2018). "Probabilistic characterization of two-dimensional soil profile by integrating CPT with MASW data." *Canadian Geotechnical Journal*, **55**(8), 1168-1181. https://doi.org/10.1139/cgj-2017-0429

Kim, E. and Hunt, R. (2017). "A public website of rock mechanics database from Earth Mechanics Institute (EMI) at Colorado School of Mines (CSM)." *Rock Mechanics and Rock Engineering*, **50**(12), 3245-3252. https://doi.org/10.1007/s00603-017-1292-1

Kıncal, C. and Koca, M.Y. (2019). "Correlations of in situ modulus of deformation with elastic modulus of intact core specimens and RMR values of andesitic rocks: A case study of the İzmir subway line." *Bulletin of Engineering Geology and the Environment*, **78**, 5281-5299. https://doi.org/10.1007/s10064-018-01443-5

Kootahi, K. and Moradi, G. (2017). "Evaluation of compression index of marine fine-grained soils by the use of index tests." *Marine Georesources and Geotechnology*, **35**(4), 548-570. https://doi.org/10.1080/1064119X.2016.1213775

Kulhawy, F.H. and Mayne, P.W. (1990). "Manual on estimating soil properties for foundation design." *Report EL-6800*, Electric Power Research Institute, Palo Alto, California. https://www.epri.com/research/products/EL-6800

Lacasse, S. and Lunne, T. (1982). "Penetration tests in two Norwegian clays." *Proceedings, 2nd European Symposium on Penetration Testing*, Amsterdam, 661-670.

Liu, S., Zou, H., Cai, G., Bheemasetti, T.V., Puppala, A.J., and Lin, J. (2016). "Multivariate correlation among resilient modulus and cone penetration test parameters of cohesive subgrade soils." *Engineering Geology*, **209**, 128-142. https://doi.org/10.1016/j.enggeo.2016.05.018

Phoon, K.K., Low, H.E., and Tan, T.S. (2004). "Variability of OCR-Q_t correlation in Singapore Upper Marine Clay." *Proceedings, Second International Conference on Geotechnical Site Characterization*, Porto, Portugal, September 19-22, 2004, **2**, 1835-1842.

Mesri, G. and Huvaj, N. (2007). "Shear strength mobilized in undrained failure of soft clay and silt deposits." *Advances in Measurement and Modeling of Soil Behavior* (GSP 173), ASCE, 1-22. https://doi.org/10.1061/40917(236)1

Mitchell, J.K. and Kopmann, J. (2013). *The Future of Geotechnical Engineering*. Report CGPR #70. Virginia Tech Center for Geotechnical Practice and Research.

Phoon, K.K. (2017). "Role of reliability calculations in geotechnical design." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **11**(1), 4-21. https://doi.org/10.1080/17499518.2016.1265653

Phoon, K.K. (2020). "The story of statistics in geotechnical engineering." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **14**(1), 3-25. https://doi.org/10.1080/17499518.2019.1700423

Phoon, K.K. and Kulhawy, F.H. (1999a). "Characterization of geotechnical variability." *Canadian Geotechnical Journal*, **36**(4), 612-624. https://doi.org/10.1139/t99-038

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

Phoon, K.K. and Ching, J. (2017). "Better correlations for geotechnical design." *A Decade of Geotechnical Advances 2008-2017*, Geotechnical Society of Singapore, Singapore, 73-102.

Phoon, K.K. and Tang, C. (2019a). "Characterization of geotechnical model uncertainty," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **13**(2), 101-130. https://doi.org/10.1080/17499518.2019.1585545

Phoon, K.K. and Tang, C. (2019b). "Effect of extrapolation on interpreted capacity and model statistics of steel H-piles." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **13**(4), 291-302. https://doi.org/10.1080/17499518.2019.1652920

Phoon, K.K., Prakoso, W.A., Wang, Y., and Ching, J. (2016). "Uncertainty representation of geotechnical design parameters," Chapter 3. *Reliability of Geotechnical Structures in ISO2394*, CRC Press/Balkema, 49-87.

Phoon, K.K., Ching, J., and Wang Y. (2019). "Managing risk in geotechnical engineering – from data to digitalization." *Proceedings, 7th International Symposium on Geotechnical Safety and Risk* (ISGSR 2019), Taipei, Taiwan, 13-34. http://rpsonline.com.sg/proceedings/isgsr2019/pdf/SL.pdf

Phoon, K.K., Ching, J., and Shuku, T. (2022a). "Challenges in data-driven site characterization." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **16**(1), 114-126. https://doi.org/10.1080/17499518.2021.1896005

Phoon, K.K., Cao, Z., Ji, J., Leung, Y.F., Najjar, S., Shuku, T., Tang, C., Yin Z-Y., Yoshida, I., and Ching, J. (2022b). "Geotechnical uncertainty, modeling, and decision making." *Soils and Foundation*, **62**(5), 101189. https://doi.org/10.1016/j.sandf.2022.101189

Phoon, K.K., Ching, J., and Cao, Z. (2022c). "Unpacking data-centric geotechnics." *Underground Space*, **7**(6), 967-989. https://doi.org/10.1016/j.undsp.2022.04.001

Phoon, K.K., Shuku, T., Ching, J., and Yoshida, I. (2022d). "Benchmark examples for data-driven site characterization." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. https://doi.org/10.1080/17499518.2022.2025541

Phoon, K.K. and Zhang, W. (2023). "Future of machine learning in geotechnics." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. https://doi.org/10.1080/17499518.2022.2087884

Prakoso, W.A. (2017). *Personal Communication*.

Prakoso, W.A. (2002). Reliability-based design of foundations on rock for transmission line and similar structure. *PhD Thesis*, Cornell University.

Sauvin, G., Vanneste, M., Vardy, M.E., Klinkvort, R.T., and Fredrik, F.C. (2019). "Machine learning and quantitative ground models for improving offshore wind site characterization." *Proceedings, Offshore Technology Conference*, Houston, Texas, OTC-29351-MS. https://doi.org/10.4043/29351-MS

Schuppener, B. (2011). "Reliability theory and safety in German geotechnical design." *Proceedings, Third International Symposium on Geotechnical Safety and Risk*, Federal Waterways Engineering and Research Institute, Germany, 527-536. https://izw.baw.de/e-medien/geotechnical-safety-and-risk-

2011/PDF/4%20Codes%20and%20Standards/4_09.pdf

Sharma, A., Ching, J., and Phoon, K.K. (2022). "A hierarchical Bayesian similarity measure for geotechnical site retrieval." *Journal of Engineering Mechanics*, ASCE, **148**(10), 04022062. https://doi.org/10.1061/(ASCE)EM.1943-7889.0002145

Shi, C. and Wang, Y. (2021). "Development of subsurface geological cross-section from limited site-specific boreholes and prior geological knowledge using iterative convolution XGBoost." *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, **147**(9), 04021082. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002583

Shuku, T., Phoon, K.K., and Yoshida, I. (2020). "Trend estimation and layer boundary detection in depth-dependent soil data using sparse Bayesian lasso." *Computers and Geotechnics*, **128**, 103845. https://doi.org/10.1016/j.compgeo.2020.103845

Shuku, T. and Phoon, K.K. (2023). "Data-driven subsurface modeling using a Markov random field model." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, under review.

Tang, C. and Phoon, K.K. (2018a). "Evaluation of model uncertainties in reliability-based design of steel H-piles in axial compression." *Canadian Geotechnical Journal*, **55**(11), 1513-1532. https://doi.org/10.1139/cgj-2017-0170

Tang, C. and Phoon, K.K. (2018b). "Statistics of model factors in reliability-based design of axially loaded driven piles in sand." *Canadian Geotechnical Journal*, **55**(11), 1592-1610. https://doi.org/10.1139/cgj-2017-0542

Tang, C. and Phoon, K.K. (2018c). "Statistics of model factors and consideration in reliability-based design of axially loaded helical piles." *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, **144**(8), 04018050. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001894

Tang, C. and Phoon, K.K. (2019a). "Evaluation of stress-dependent methods for the punch-through capacity of foundations in clay with sand." *ASCE-ASME Journal of Risk Uncertainty in Engineering Systems, Part A: Civil Engineering*, **5**(3), 04019008. https://doi.org/10.1061/AJRUA6.0001016

Tang, C. and Phoon, K.K. (2019b). "Characterization of model uncertainty in predicting axial resistance of piles driven into clay." *Canadian Geotechnical Journal*, **56**(8), 1098-1118. https://doi.org/10.1139/cgj-2018-0386

Tang, C. and Phoon, K.K. (2020). "Statistical evaluation of model factors in reliability calibration of high-displacement helical piles under axial loading." *Canadian Geotechnical Journal*, **57**(2), 246-262. https://doi.org/10.1139/cgj-2018-0754

Tang, C. and Phoon, K.K. (2021). *Model Uncertainties in Foundation Design*. CRC Press. https://doi.org/10.1201/9780429024993

Tang, C., Phoon, K.K., and Chen, Y.J. (2019). "Statistical analyses of model factors in reliability-based limit state design of drilled shafts under axial loading." *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, **145**(9), 05019042. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002087

Tang, C., Phoon, K.K., Li, D.Q., and Akbas, S.O. (2020). "Expanded database assessment of design methods for spread foundations under axial compression and uplift loading." *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, **146**(11), 04020119. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002373

Tao, Y., Phoon, K.K., Sun, H., and Cai, Y. (2023). "Physics-informed hierarchical Bayesian model for predicting small-strain stiffness of sand." *Canadian Geotechnical Journal*, under review.

Wang, H., Wang, X., Wellmann, J.F., and Liang, R.Y. (2019). "A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data." *Canadian Geotechnical Journal*, **56**(8), 1184-1205. https://doi.org/10.1139/cgj-2017-0709

Wang, Y., Hu, Y., and Phoon, K.K. (2021). "Non-parametric modelling and simulation of spatiotemporally varying geo-data." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, **16**(1), 77-97. https://doi.org/10.1080/17499518.2021.1971258

Wu, S., Ching, J., and Phoon, K.K. (2022). "Quasi-site-specific soil property prediction using a cluster-based hierarchical Bayesian model." *Structural Safety*, **99**, 102253. https://doi.org/10.1016/j.strusafe.2022.102253

Xiao, S., Zhang, J., Ye, J., and Zheng, J. (2021a). "Establishing region-specific N–Vs relationships through hierarchical Bayesian modeling." *Engineering Geology*, **287**, 106105. https://doi.org/10.1016/j.enggeo.2021.106105

Xiao, T., Zou, H.F., Yin, K.S., Du, Y., and Zhang, L.M. (2021b). "Machine learning-enhanced soil classification by integrating borehole and CPTU data with noise filtering." *Bulletin of Engineering Geology and the Environment*, **80**, 9157-9171. https://doi.org/10.1007/s10064-021-02478-x

Xu, D. and Tian, Y. (2015). "A comprehensive survey of clustering algorithms." *Annals of Data Science*, **2**, 165-193. https://doi.org/10.1007/s40745-015-0040-1

Xu, J., Wang, Y., and Zhang, L. (2021). "Interpolation of extremely sparse geo-data by data fusion and collaborative Bayesian compressive sampling." *Computers and Geotechnics*, **134**, 104098. https://doi.org/10.1016/j.compgeo.2021.104098

Yoshida, I., Tomizawa, Y., and Otake, Y. (2021). "Estimation of trend and random components of conditional random field using gaussian process regression." *Computers and Geotechnics*, **136**, 104179. https://doi.org/10.1016/j.compgeo.2021.104179

Zhang, D.M., Zhou, Y., Phoon, K.K., and Huang, H.W. (2020). "Multivariate probability distribution of Shanghai clay properties. *Engineering Geology*, **273**, 105675. https://doi.org/10.1016/j.enggeo.2020.105675

Zou, H., Liu, S., Cai, G., Puppala, A.J., and Bheemasetti, T.V. (2017). "Multivariate correlation analysis of seismic piezocone penetration (SCPTU) parameters and design properties of Jiangsu quaternary cohesive soils." *Engineering Geology*, **228**, 11-38. https://doi.org/10.1016/j.enggeo.2017.07.005