Probabilistic Engineering Mechanics 71 (2023) 103379

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Model calibration: A hierarchical Bayesian approach

Nezam Bozorgzadeh *,[1], Zhongqiang Liu *,[2], Farrokh Nadim[3], Suzanne Lacasse[4]

Norwegian Geotechnical Institute, PO Box 3930, Ullevaal St., NO-0806, Oslo, Norway

ARTICLE INFO

Keywords:
Bayesian hierarchical regression
Method bias
Parameter uncertainty
Site-specific data
Axial capacity of piles
CPT design method

ABSTRACT

Many geotechnical engineering models are empirical and calibrated based on data gathered from various sites/projects, using optimisation algorithms with criteria like least squared errors or minimising the coefficient of variation of method bias with the constraint of mean bias equal to unity. This paper discusses the use of hierarchical Bayesian regression models for the same purpose. A database of axial capacity of piles in predominantly clay sites and a CPT-based design model, compiled and developed as part of a Joint Industry Project (JIP) led by the Norwegian Geotechnical Institute (NGI), is used for demonstration. The analyses focus on two related areas that the traditional approaches overlook: (i) quantification of uncertainty in the estimated parameters of the model, and (ii) modelling site-dependency of the model parameters (i.e., between-group variation). The former is important in the context of reliability-based design and contributes to establishing confidence in estimated reliability indices, particularly when only limited data are available. The latter expands our understanding regarding the domain of applicability of a model; that is, if a model is broadly applicable or highly site-dependent. The benefits of the proposed Bayesian approach are highlighted with a prediction exercise where the calibrated models are used in conjunction with limited site or project-specific data.

1. Introduction

Calibration of (empirical) geotechnical models involves gathering data from different sources that are assumed similar to each other in the sense that they fall within an envelope that is representative of the behaviour or application being investigated. Such data compilations are sometimes referred to as generic databases. Then, the corresponding "best-fit" values for the parameters of the model are found.

The best-fit parameters are often the outcome of an optimisation exercise formulated either as least-squares regression analysis or by minimising the coefficient of variation (COV) of method bias (usually defined as the ratio of measured to calculated values; also referred to as model factor, model bias; hereafter, bias for brevity) with the constraint of mean bias equal to unity. A few examples of such calibrations are [1–4], and [5].

Regardless of the methodology used, many model calibrations reported in the geotechnical engineering literature have shortcomings in two aspects. First, uncertainty in the best-fit coefficients, i.e. statistical parameter uncertainty, is overlooked. That is, only point estimates for coefficients are reported with little or no attention to the precision of such estimates. Reporting and propagating statistical parameter uncertainty is particularly important when dealing with geotechnical data where variations are typically large and the data limited. In the context of reliability-based design (RBD) and load and resistance factor design (LRFD) calibration, acknowledging this statistical parameter uncertainty results in uncertain estimates of reliability indices or resistance factors, as explored by Surles et al. [6] using Bootstrapping and Bozorgzadeh and Bathurst [7] from a Bayesian perspective.

Second, more often than not, the generic databases gathered for purposes such as model calibration, RBD or LRFD calibration are structured in that they could be parsed into sub-groups (e.g. by project, region, site, location within site, laboratory and such) that are similar to each other in the sense that they fall under a representative envelope for the application being investigated, but are not necessarily identical conditions. Therefore, there could be quantifiable variations between such sub-groups (e.g. [8,9]). Typical calibration of geotechnical models ignores such structures in the data and assumes the model parameters to be representative of all the data. In other words, the model parameters for all possible sub-groups are assumed to be identical and estimated from all the data. This could be misleading, or at the least sub-optimal, because all the residual variation is expressed as a lumped variance, portions of which could potentially be attributed to the existence of

* Corresponding authors.
E-mail addresses: Nezam.Bozorgzadeh@ngi.no (N. Bozorgzadeh), Zhongqiang.Liu@ngi.no (Z. Liu), Farrokh.Nadim@ngi.no (F. Nadim), Suzanne.Lacasse@ngi.no (S. Lacasse).
[1] Project Engineer.
[2] Senior Engineer.
[3] Technical Director.
[4] Expert Adviser.

https://doi.org/10.1016/j.probengmech.2022.103379
Received 12 October 2022; Accepted 2 November 2022
Available online 9 November 2022

the above-mentioned sub-groups. Hierarchical Bayesian models provide a means of relaxing the assumption of identical parameters, allowing for model parameters to vary from each other but still learn from one another when there are similarities among them (e.g. [10]). In other words, each group-specific parameter is estimated from the data within that group, but also "borrows strength" (or precision) from the data in all other groups [11].

The underlying ideas and theoretical basis for hierarchical modelling date back to the 1950s and 1960s, with applied Bayesian analysis using hierarchical models gaining some popularity in the 1980s and 1990s and becoming more common afterwards with the availability of modern computers and Markov Chain Monte Carlo (MCMC) algorithms. A review of the theoretical and applied advancements in hierarchical modelling is beyond the scope of this paper but a comprehensive overview may be found in the bibliographic note in Chapter 5 of [10].

Interestingly, hierarchical models are absent as a general approach to data analysis from the decades-long history of application of Bayesian techniques in geotechnical engineering (e.g. see review papers by Zhang et al. [12] and Baecher [13]). Zhang et al. [14,15] are among early examples of using hierarchical models for systematic modelling of between-site variations. Bozorgzadeh et al. [16] and Bozorgzadeh and Bathurst [8] provide relatively accessible introductions to hierarchical modelling as a general means of approaching geotechnical engineering data. They discuss the relevance of hierarchical modelling for geotechnical applications to be three-fold: first, they can be seen as tools for understanding the variations in the data (analysis of variance); second, they can be used to make predictions about a yet-to-be-observed subgroup, which is the design problem in a future project/site; and third, they provide improved (partially pooled, see Section 3) estimates of a group-specific parameter for which some group-specific data are also be available. An instance of the latter is the case where limited site or project-specific data are available, and it is desirable to combine this site-specific information with a generic database. Ching and Phoon [17] and Ching et al. [9,18] are notable examples of application of hierarchical models for predicting geotechnical parameters from empirical relationships based on Cone Penetration Test (CPT) data.

With axial capacity of piles in clay as an example, this paper explores calibrating geotechnical models using hierarchical Bayesian regression while discussing the above-raised two issues of quantifying statistical parameter uncertainty and between-group variations. The paper is organised as follows. Section 2 introduces the data and design model for the axial capacity of piles in clay. Section 3 presents the regression modelling including (statistical) model checking and comparison. Finally, Section 4 discusses some of the benefits of hierarchical modelling via selected examples of predicting site-specific data.

2. Axial pile capacity formulation and database of pile load tests

With the increasing application of the CPT in recent years, there has been a parallel increase in interest in direct CPT-based geotechnical design models. The new Unified, unaffiliated CPT-based method for axial pile capacity calculation [19] is a recent example of such models. This model was the outcome of two Joint Industry Projects (JIP) under the management of the Norwegian Geotechnical Institute (NGI), which led to the creation of two "Unified" databases of high-quality instrumented pile load tests, one in predominantly sand and the other in predominantly clay sites (hereafter sand and clay for brevity) [20]. The Unified database was established with the consensus of university representatives in the profession and personnel in multiple companies from the offshore energy sector. Models for the design of axial capacity of piles in sand and clay were developed. The following provides the background information about the pile load test data and the model for axial capacity of piles in clay that is necessary for formulating the model calibration regression problem in this paper. Details beyond what is presented here may be found in [21,22] and [23].

2.1. Unified database

A full description of the "Unified" database in clay is provided by Lehane et al. [20]. A total of 49 pile load tests were selected after careful screening based on stringent selection criteria described by Lehane et al. [20]. The CPT data for 10 tests close to or within Zone 1 of the soil behaviour type (SBT) chart [24] indicated sensitive clay. The statistical analyses focused on the clays outside of Zone 1, so 39 pile loads from 20 sites. An overview of the pile load tests in clay in the Unified database is provided in Table 1, with more details on individual pile configurations, end conditions, loading direction, equalisation time, and measured axial capacity presented in Table A.1 in Appendix.

2.2. Axial pile capacity model

The axial capacity of piles consists of contributions from shaft skin friction and end bearing. The Unified, unaffiliated model for the skin friction component of the axial capacity of piles in clay is [22,23]:

$$\tau_f = A q_t \left[\max\left(\frac{h}{D^*}, 1\right)\right]^{-c} \tag{1}$$

where τ_f is the total (clay) skin friction, q_t is the corrected cone resistance, h is the distance from pile tip to a horizon, and D^* is the equivalent pile diameter calculated as $(D^2 - D_i^2)^{0.5}$ for open-ended piles, and reduces to D for closed ended piles; D is outer pile diameter and D_i is inner pile diameter, reflecting the lower level of displacement induced during installation of a pipe pile. The term h/D^* represents length effect. A and c are coefficients to be determined.

The end bearing of compression piles in clay usually represents a small fraction of the total capacity. The Unified unaffiliated model assumes the end bearing at a pile movement of 10% of the pile diameter $q_{b0.1}$ to be [22,23]:

$$q_{b0.1} = 0.8 q_t \quad \text{(closed-ended pile)} \tag{2a}$$

$$q_{b0.1} = 0.4 q_t \quad \text{(open-ended pile)} \tag{2b}$$

It should be noted that the Unified unaffiliated model recommends using the essentially equivalent approximation for end bearing:

$$q_{b0.1} = q_t[0.2 + 0.6(D^*/D)^2] \tag{2c}$$

For the calibration example in this paper, the following assumptions are made for simplicity. First, the end bearing is taken as deterministic. Second, the piles in the database are driven in *dominantly* clay layers; the contribution of sand/silt layers (if present) is calculated using the Unified CPT-based method in sand [19], also taken to be deterministic. Therefore, the measured skin friction capacity for calibrating the model in Eq. (1) is calculated as the total measured capacity minus the contributions of sand/silt layers and end bearing.

Hereafter, the term "model", when used in isolation, refers to a statistical model. To prevent confusion between statistical models and the geotechnical model, the latter is explicitly referred to as the pile capacity or skin friction model.

3. Statistical analysis of data

3.1. Regression models

The coefficients of the model in Eq. (1) are estimated using Bayesian regression. Two general regression modelling approaches can be adopted when calibrating such a model to be used for prediction and design: *complete pooling* and *partial pooling*. Below we briefly introduce each approach and the corresponding regression models. More detailed discussion of these modelling approaches may be found in seminal Bayesian textbooks such as [11] and [10], with summary overview and discussions in the context of geotechnical data in e.g. [16].

Table 1
Overview of pile load tests in clay in the unified database.

Characteristics of piles tested		Number of pile load tests		
		Closed	Open	All
Material	Steel	13	15	28
	Concrete	7	4	11
Shape	Circular	18	19	37
	Square (and others)	2	0	2
Loading direction	Compression	10 (CEC)	11 (OEC)	21 (7 with $Q_{b,ult}$ data)
	Tension	10 (CET)	8 (OET)	18
Total		20	19	39

CEC: Closed-ended pile tested in compression; OEC: Open-ended pile tested in compression; CET: Closed-ended pile tested in tension; OET: Open-ended pile tested in tension; $Q_{b,ult}$: Measured ultimate end bearing capacity.

3.1.1. Complete pooling

As mentioned earlier, the common approach to model calibration in geotechnical engineering overlooks the possibility that the available data could be parsed into multiple sub-groups (here, pile load tests from different sites): all the data are pooled into a single large group from which the model parameters are estimated (hence complete pooling). This model is also referred to as the *identical parameter* model because it assumes that the model parameters for the sub-groups are identical. Here, a Bayesian regression model is formulated on the log–log scale; for $i = 1, 2, 3, \ldots, N$ pile load tests:

$$\mathrm{Ln}(\tau_{f_{[i]}}) = \mathrm{Ln}(A) + \mathrm{Ln}(q_{t_{[i]}}) - c \cdot \mathrm{Ln}(\left[\max\left(\frac{h_{[i]}}{D^*_{[i]}}, 1\right)\right]) + e_{[i]} \tag{3a}$$

$$e_{[i]} \sim \mathrm{Normal}(0, \sigma) \tag{3b}$$

where σ is the standard deviation of $\mathrm{Ln}(\tau_f)$. In other words, Eq. (3a) describes a model where the natural logarithm of total skin friction is normally distributed with mean equal to the natural logarithm of the right-hand-side of Eq. (1) and standard deviation σ. The following priors are assigned to A, c and σ:

$$A \sim \mathrm{HalfNormal}(0, 1) \tag{3c}$$

$$c \sim \mathrm{HalfNormal}(0, 1) \tag{3d}$$

$$\sigma \sim \mathrm{HalfNormal}(0, 1) \tag{3e}$$

Regarding the above choice of prior distributions, it is noted that statistical analyses of geotechnical data in the literature usually adopt default flat priors such as Uniform(0, U) (with U chosen to be a large value such as 100 or 1000 depending on the scale of the data) or a normal distribution with zero mean and large standard deviation. However, more recent recommendations on choosing prior distributions (e.g. [25]) point out that such traditional default flat priors assign non-zero probabilities to extreme parameter values that are known to be unreasonable. Also, priors with hard boundaries are recommended only when justified by mathematical or physical constraints, e.g. standard deviation must be positive. The half-normal is recommended as a generic weakly informative prior especially for situations where the (scaled) parameters are roughly on unit scale; half-t or half-Cauchy are recommended when heavy tails are expected. In the context of the current study, the parameters A and c are expected to be between 0 and 1, with previous similar work on piles in predominantly sand layers ([20]) suggesting that the smaller values in that range are more likely. The measured pile capacity data (Table A.1) span about 5 units on the log-scale which roughly corresponds to a standard deviation of 1.25 (assuming approximate normality). However, the residual standard deviation of the regression model is expected to be much smaller than the standard deviation of the unmodelled observations because the model should capture considerable portions of the raw variation in the data. The HalfNormal$^+$(0, 1) is bounded to be positive, has a mean of about 0.77, median of about 0.63, and 95%-th percentile of 1.95, so it excludes (or more precisely, assign practically zero prior probability to) extreme parameter values that we know to be unreasonable. It also assigns roughly 0.7 probability to values smaller than 1. The assigned HalfNormal priors express soft constraints on the parameter space and contribute to more efficient Markov Chain Monte Carlo (MCMC) sampling without distorting posterior inferences; overall, the posterior means obtained from the model of Eq. (3) are comparable to best-fit values from an optimisation analysis where the COV of the bias is minimised with the constraint of mean bias equal to unity [23]. Finally, it is worth mentioning that fitting the model using Uniform(0, 1) priors results in essentially the same (within MC error) posteriors.

3.1.2. Partial pooling (hierarchical model)

In Bayesian hierarchical modelling, statistical parameters of different sub-groups in the data are assumed to be similar but not identical, which is an informal expression of the mathematical concept of “exchangeable” parameters. Exchangeable parameters can be thought of as being drawn from a population distribution, and thus can be assigned a common prior distribution with unknown parameters ([11]). In other words, a statistical model is assumed for the parameters of the data model, the parameters of which (i.e. the hyper-parameters) would be estimated from the data. Therefore, each group-specific parameter is estimated from two sources of information: (i) directly from the data within the corresponding group, and (ii) from the data in all other groups, but indirectly and through the common prior distribution.

From a geotechnical engineering perspective, the ideal hierarchical model for axial capacity of piles is one with hierarchical structures specified for coefficients A and c, as well as standard deviation σ. That is, a model with site-specific As, site-specific cs and site-specific σs, all of which are estimated from within-site data but also borrow information from the data belonging to other sites via some common prior distribution. However, the available data ($N = 39$ from $J = 20$ sites, with many sites comprising very few observations) do not allow meaningful estimation of such a complex model, and thus a simpler hierarchical model should be considered.

The small number of observations in each site means that the within-site data contribute very little to estimation of a site-specific residual standard deviation σ. In case of a model with hierarchical standard deviation σ, the group-specific standard deviations will essentially be equal to the higher-level population standard deviation, which itself is being estimated from the limited $J = 20$ observations and thus potentially noisy. Therefore, assuming a model with common standard deviation does not seem unreasonable. Between the two parameters of the axial capacity model, A is directly multiplied by the corrected cone resistance q_t, which is the key parameter representing soil strength in the model and derived from CPT data. Coefficient c represents the rate of friction fatigue with distance from the pile tip. Therefore, between A and c, A is judged to be the better representative of site conditions. Thus, a regression model with hierarchical A (varying by site) and common c and σ is formulated for $j = 1, 2, \ldots, 20$ sites:

$$\mathrm{Ln}(\tau_{f_{[i,j]}}) = \mathrm{Ln}\left(A_{[j]}\right) + \mathrm{Ln}(q_{t_{[i,j]}}) - c \cdot \mathrm{Ln}(\left[\max\left(\frac{h_{[i,j]}}{D^*_{[i,j]}}, 1\right)\right]) + e_{[i]} \tag{4a}$$

$$e_{[i]} \sim \mathrm{Normal}(0, \sigma) \tag{4b}$$

$$A_{[j]} \sim \text{Lognormal}(\mu_A, \sigma_A) \tag{4c}$$

$$c \sim \text{HalfNormal}(0, 1) \tag{4d}$$

$$\sigma \sim \text{HalfNormal}(0, 1) \tag{4e}$$

$$\mu_A \sim \text{Normal}(0, 1) \tag{4f}$$

$$\sigma_A \sim \text{HalfNormal}(0, 1) \tag{4g}$$

Eq. (4c) states that the coefficients A have a common distribution with unknown parameters. In other words, the A of each site has been drawn from a population distribution with mean μ_A and standard deviation σ_A. It is through this population distribution that when learning about A of a site from data belonging to that site we also "borrow strength" from the data from all the other sites.

The parameter σ_A is the standard deviation among the (natural logarithm of the) site-specific As and it is instructive to explore its limits. If $\sigma_A \to 0$, then there is no variation among groups, i.e. all the $A_{[j]}$s would be equal; this is the complete pooling model. On the other extreme, if $\sigma_A \to \infty$, then no common distribution of As exists, i.e. a *no pooling* model; the As are independent. Hierarchical estimates are close to complete pooling estimates if σ_A is small (relative to the total variation in the data) and close to the no pooling estimates if σ_A is large. Therefore, it can be seen that hierarchical models are the more general formulation of the problem, with the no pooling and complete pooling models as their limiting cases.

It is worth mentioning that the above-mentioned *no pooling* model (site-specific As, common c and σ) can be formulated by assigning to each $A_{[j]}$ an independent vague prior (as opposed to the common prior in the hierarchical model). No pooling models with non-informative priors are rarely of interest for analysing geotechnical data because the usually limited quantity of project or site-specific data do not allow reliable parameter estimations and it is necessary to borrow information from elsewhere. Also, by definition, they cannot be used for predictions beyond the data they have been fitted to.

3.2. Results

3.2.1. Posterior distributions

All regression models are fitted using the computer software Stan [26] with pre- and post-processing of the data and results in R [27]. The Markov Chain Monte Carlo (MCMC) consisted of running three parallel chains, each with 6000 total simulations and 1000 warm-up iterations (15000 simulations in total). All chains were examined for lack of convergence both visually and using quantitative diagnostics.

Table 2 provides posterior summary statistics for the complete pooling and partial pooling (hierarchical) models. The mean of a posterior distribution can be thought of as a point estimate and is comparable to the best-fit value obtained from a least-squares regression or minimising bias COV. The spread of posterior distributions (expressed as the standard deviation and 95% credible interval (CI)) emphasises the uncertainty associated with the parameters.

Fig. 1 shows the joint posterior distribution of A and c from the two models. Note that in this figure, A from the hierarchical model is the population level A, i.e. posterior simulations from Lognormal(μ_A, σ_A), which accounts for posterior uncertainty in the hyper-parameters μ_A and σ_A, but more importantly for between-site variation, i.e. the deviations of individual site-specific A values from the population-level mean A . Fig. 2 shows a visual summary of posterior distributions of site-specific As and how they vary from one-another and how they compare to the complete pooling A. The hierarchical estimate of A (for a new site) is more uncertain than that of the complete pooling. However, the estimate for residual standard deviation is smaller in the hierarchical model. In other words, and informally, some of the residual standard deviation in the complete pooling model ($\sigma = 0.24$) has been *explained* by the variation between As of different sites. The 95% CI for each A emphasises uncertainty in the estimated value of A for the corresponding site.

Table 2
Posterior summary statistics for the complete and partial pooling models.

Model	Parameter	Posterior	
		Mean$_{(SD)}$ [a]	95% CI [b]
Complete pooling	A	$0.06_{(0.01)}$	(0.04, 0.08)
	c	$0.19_{(0.06)}$	(0.07, 0.31)
	σ	$0.24_{(0.03)}$	(0.19, 0.29)
Partial pooling	μ_A	$-2.84_{(0.15)}$	(−3.15, −2.54)
	σ_A	$0.22_{(0.05)}$	(0.14, 0.32)
	m_A [c]	$0.06_{(0.01)}$	(0.04, 0.08)
	s_A [d]	$0.01_{(0.00)}$	(0.01, 0.02)
	c	$0.19_{(0.05)}$	(0.09, 0.29)
	σ	$0.12_{(0.02)}$	(0.08, 0.17)

[a] Standard deviation.
[b] Credible interval.
[c] Mean of A.
[d] Standard deviation of A.

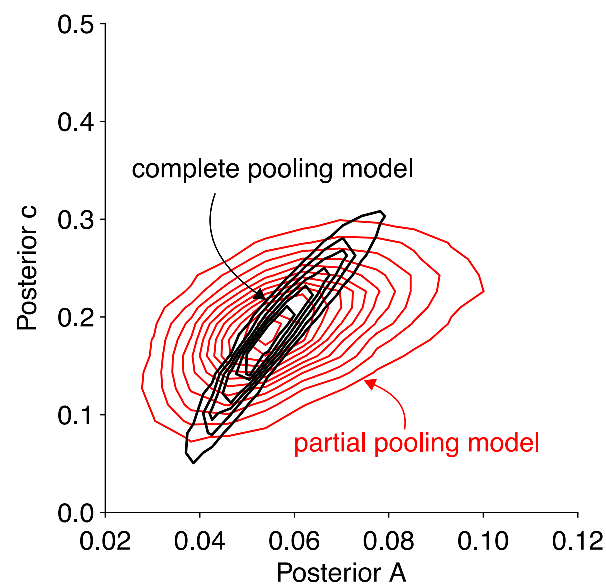

Fig. 1. Joint posterior distribution of A and c (iso-density contours).

Comparing the measured skin friction with the fitted values from the two models in Fig. 3 is another visualisation of the effect of hierarchical modelling; for better visualisation, these are presented on the log–log scale. First, it is noted that the fitted values have posterior distributions, summarised in the figure as point estimates accompanied with 95% CIs. At each MCMC iteration a value of A and a value of c are sampled; these can be used in conjunction with the q_t profile and pile dimension data to calculate a value for pile capacity implied by the model, i.e. the fitted value. Repeating this for all MCMC iterations results in simulations from the posterior distribution of fitted values that accounts for uncertainty in the parameters of the model, i.e. A and c. The estimates from the hierarchical model are generally closer to the one-to-one line, demonstrating the higher influence of site-specific data on these estimates. However, the hierarchical estimates also show generally wider CIs relative to the complete pooling estimates. This is the hierarchical model emphasising the uncertainty in site-specific estimates with few data points. In other words, the hierarchical model adapts to the site-specific data but does not overfit to these limited data.

3.2.2. Model checking and comparison

Fig. 4 shows the standardised residuals (i.e. the difference between the observed and fitted values, divided by the standard deviation of the regression) of the complete pooling and partial pooling regression models. The standardised residuals are expected to generally lie between

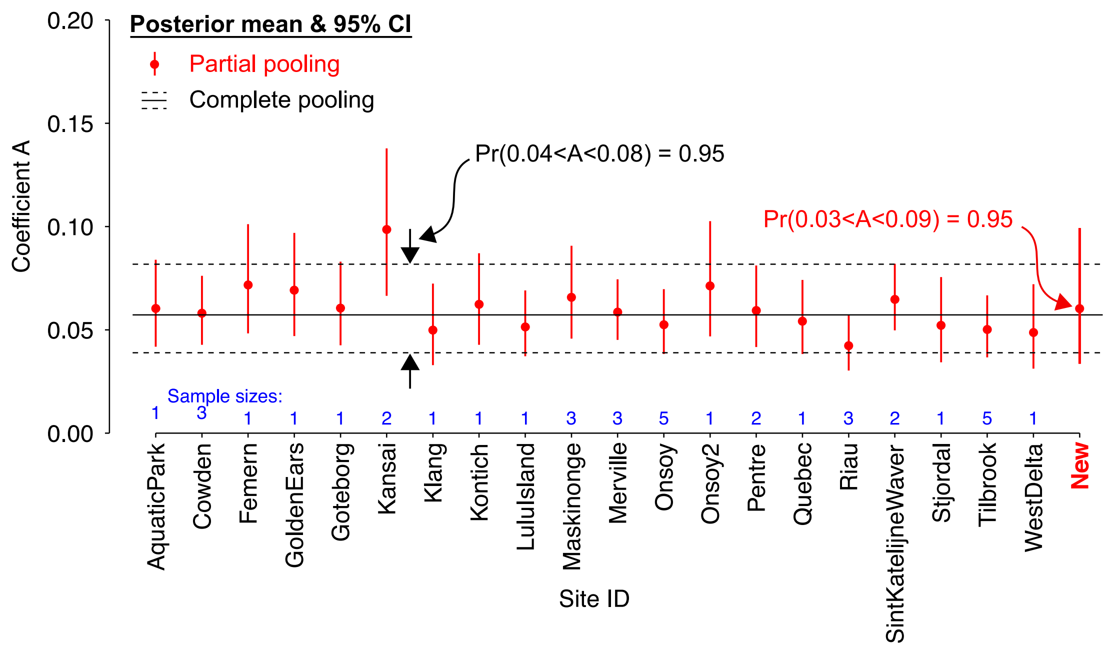

Fig. 2. Posterior distribution of A from complete and partial pooling models.

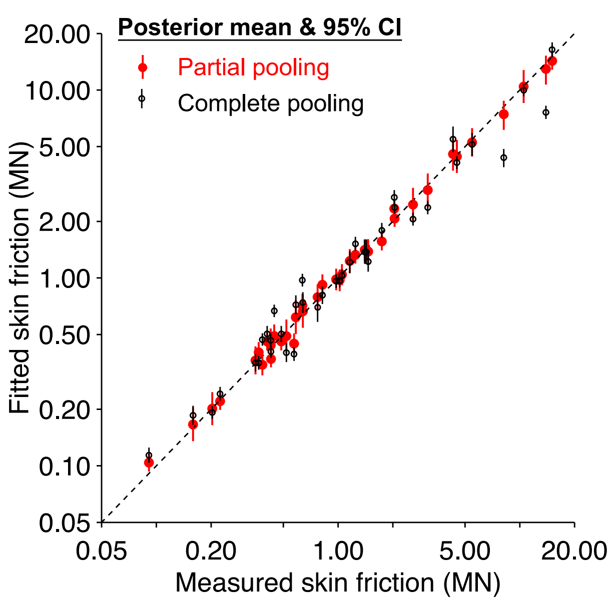

Fig. 3. Fitted vs. measured skin friction from complete and partial pooling models.

−1.96 and +1.96 (95% of the area under the standard normal probability density function lies in this range). It is noted that the residuals have posterior uncertainty because the fitted values are calculated using uncertain model parameters A and c. From the complete pooling model, residuals for two (Kansai bridge, see Table A.1) observations (out of the total 39 data points) fall outside of this 95% range. Furthermore, the probability of observing a value as or more extreme as each of these two observations is less than about 0.01, indicating that these observations are very unlikely according to the complete pooling model, i.e., they are potential outliers. Removing both of them and fitting the complete pooling model with $N = 37$ data points results in a posterior standard deviation with mean of about 0.19 (compare to estimate of 0.24 from all data in Table 1).

Currently, the underlying reason for these rather extreme observations has not been identified, but some suspected contributing factors are potentially higher uncertainties in end bearing capacity interpretations and thus poorly interpreted shaft fictions, and the fact that the two 1.5 m diameter piles at Kansai were driven into a clay deposit that included significant sand layers. The complete pooling model underpredicts the capacity of Kansai piles so its application is conservative from an engineering design point of view. Nevertheless, the residuals and tail probabilities for these observations suggest that the extent of applicability of the complete pooling model to piles with larger diameters requires further investigation.

The hierarchical model generally fits better. The two Kansai observations are predicted well. There is one data point (data point number 36, from the Maskinongé site, see Table A.1) with a posterior mean residual that falls close to +1.96. Nevertheless, this posterior residual is considerably uncertain, with its left (lower) tail extending well below +1.96. Furthermore, the associated posterior predictive tail probabilities for this observation is about 0.03, and thus not exhibiting worrying signs of misfit.

The predictive accuracy of the complete and partial pooling models can be compared using the "leave-one-out" information criterion (LOOIC). LOOIC uses log pointwise predictive density (lppd) as a measure of goodness-of-fit, in conjunction with leave-one-out cross-validation (LOOCV) [10,28]. Smaller LOOIC indicates higher predictive accuracy. LOOIC estimates for the complete and partial pooling models are $0.36_{(11.50)}$ and $-33.12_{(8.50)}$ (subscripts are standard errors). The difference in LOOIC is $33.48_{(12.38)}$ in favour of the partial pooling model. This difference in LOOIC is large and about 2.7 standard errors away from zero. Nevertheless, it is generally recognised that LOOIC standard errors could be underestimated, therefore the complete pooling model should not necessarily be ruled out as an incorrect model; rather, its application should be justified in the context that it is being used. Finally, it is worth mentioning that although formulating and fitting hierarchical models is more involved, we follow Gelman et al. [29] who discuss that hierarchical models are generally preferable to complete and no pooling models: first, we rarely believe that between-group variations are truly zero, and second, as discussed earlier, in case of this variation being relatively small/large the hierarchical model gives estimates close to the no or complete pooling models.

3.2.3. Posterior method bias

It is common in geotechnical engineering applications to quantify the performance of a model using method bias λ, i.e. the ratio between

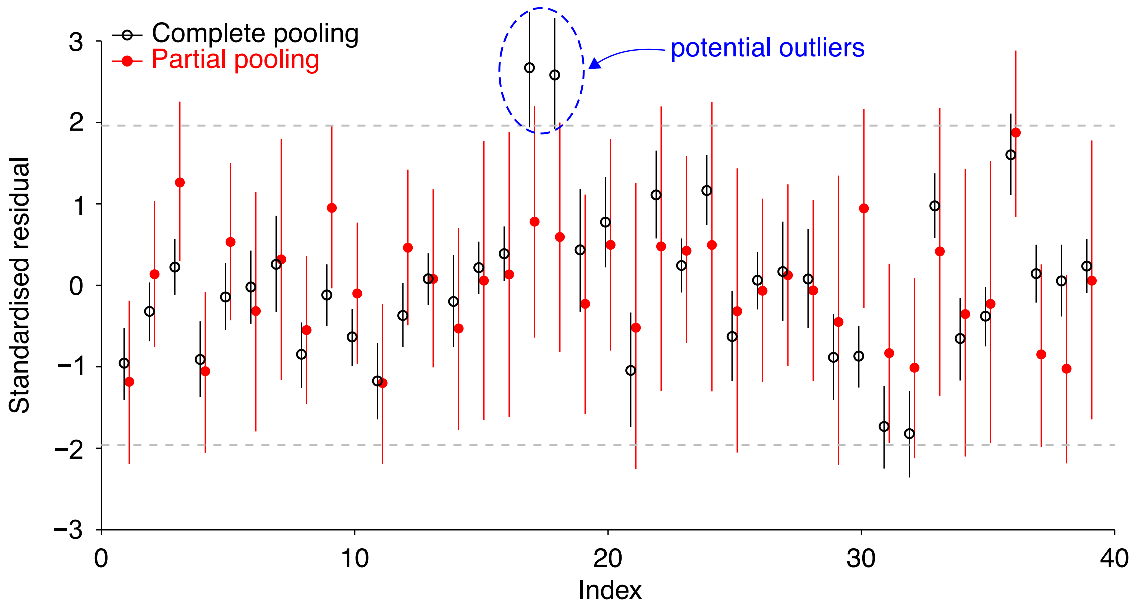

Fig. 4. Residuals from the complete and partial pooling models.

Table 3
Posterior summary statistics for method bias.

Model	Parameter	Posterior	
		Mean$_{(SD*)}$	95% CI**
Complete pooling	Bias mean	1.00$_{(0.04)}$	(0.92, 1.08)
	Bias COV	0.25$_{(0.01)}$	(0.246, 0.280)
Partial pooling	Bias mean	1.00$_{(0.02)}$	(0.97, 1.04)
	Bias COV	0.11(0.01)	(0.09, 0.15)

the measured and calculated values of a quantity of interest X: $\lambda_X = X_m/X_c$. Mean method bias of unity implies an on average accurate model; the smaller the method bias COV the more precise the model. Here, the fitted values from the regression model can be used to obtain method bias. It is recognised that the errors of the regression model fitted on the log-scale are normally distributed. It follows that logarithm of bias is normally distributed which in turn means that bias is lognormally distributed. The mean and COV of this lognormal distribution of bias can be simulated from the posterior distributions of the regression model.

At each MCMC iteration, values of A and c are simulated, from which the calculated (fitted) value of skin friction can be simulated for each pile as discussed earlier in the context of Fig. 3. The calculated values can in turn be used in conjunction with the corresponding measured values to obtain bias values for each data point. Repeating this at all MCMC iterations results in posterior distributions of bias for each pile. Fig. 5 summarises the posterior distributions of the $N = 39$ biases. Note that there is posterior uncertainty associated with each individual bias due to the posterior uncertainty in the model parameters. As mentioned earlier, more often than not, this uncertainty is overlooked in the geotechnical literature. Finally, mean and COV of method bias can also be simulated. At each MCMC iteration, 39 bias values are simulated, the mean and COV of which can be calculated. Repeating this at all MCMC iterations results in posterior distributions of bias mean and COV. These are shown in Fig. 6 with summary statistics reported in Table 3. The histograms of bias mean and COV emphasise that there is uncertainty associated with these statistics. It should be noted that bias COV represents the residual variation and has a role analogous to the parameter σ in the regression models. More details about bias and variability may be found in e.g. [30].

The general pattern in the above results is that individual bias estimates from the partial pooling model exhibit larger uncertainty. On the other hand, the variation among these (informally, the fluctuation of estimates from one data point to another) is smaller compared to the complete pooling model biases. The same can be concluded from the posterior distributions of the COV of bias in Fig. 6b.

The two models result in different estimates of parameter uncertainty and residual variation. Generally, the COV quantifies the residual (i.e. un-modelled) variation in the data which is sometimes referred to as aleatory uncertainty, while the spread in the posterior distributions of bias mean and COV quantifies uncertainty in what the values of these parameters should be which is referred to as epistemic uncertainty. It is crucial to mention that these epistemic and aleatory uncertainties are not properties of the outside world but depend on the models used to generalise from observed samples to populations (e.g. [31,32]). For instance, the hierarchical model in this paper explains some of the variation in the data to be variation between sites, hence giving on average smaller bias COV (stemming from residual variation) compared to the *status quo* complete pooling model. However, the distribution of bias COV is wider for the partial pooling model. Both models indicate that bias mean is close to unity, but hierarchical model conveys this with more certainty. This is reasonable since the individual hierarchical estimates of bias are more influenced by the corresponding observed values and hence closer to unity.

The next section demonstrates the importance of recognising and quantifying these two types of uncertainty in the context of updating a model with site-specific data.

4. Model predictions and updating

As mentioned earlier, the possibility of augmenting (limited) site-specific data with a generic database is one of the reasons that makes hierarchical modelling attractive for geotechnical applications. For instance, Bozorgzadeh et al. [16] discuss how hierarchical modelling should be used as a means of constructing informative geotechnical prior distributions. To explore this further, we discuss predictions from the two models from a leave-one-out analysis. From the total $N = 39$ observations, a selected data point is excluded from the analysis, the two models are then fitted to the remaining 38 data points, and predictions are simulated from the posteriors of each model for the left-out data point. These are essentially the prior predictive for the left-out data. In the next step, the models are updated by observing the left-out pile load test. Below we first provide some general details about how to simulate pile capacity for a new site with no pile load measurement, and then four selected examples of updating are discussed.

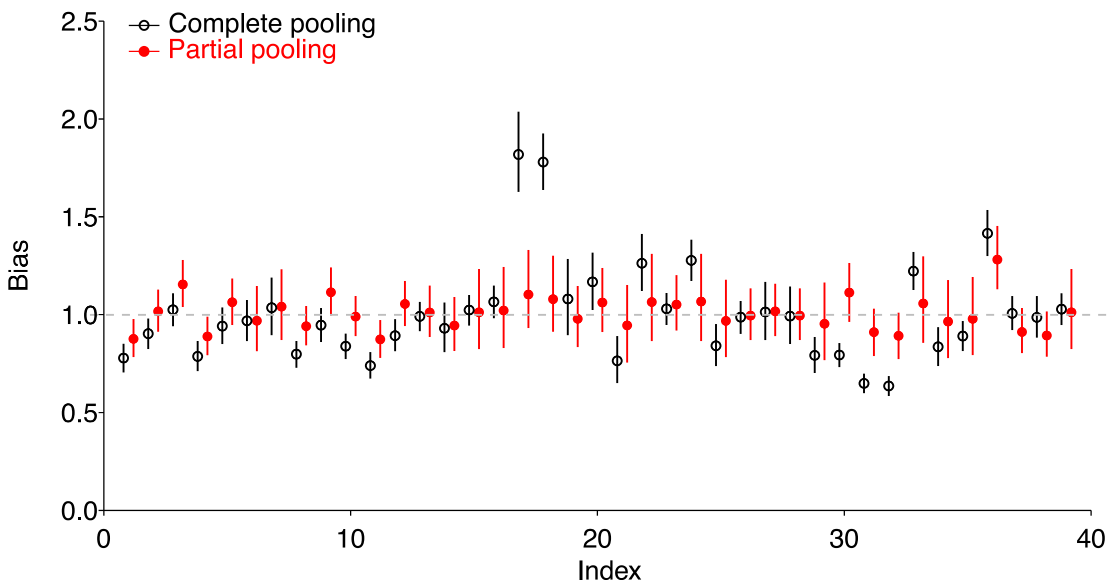

Fig. 5. Posterior mean and 95% CI for individual method bias values.

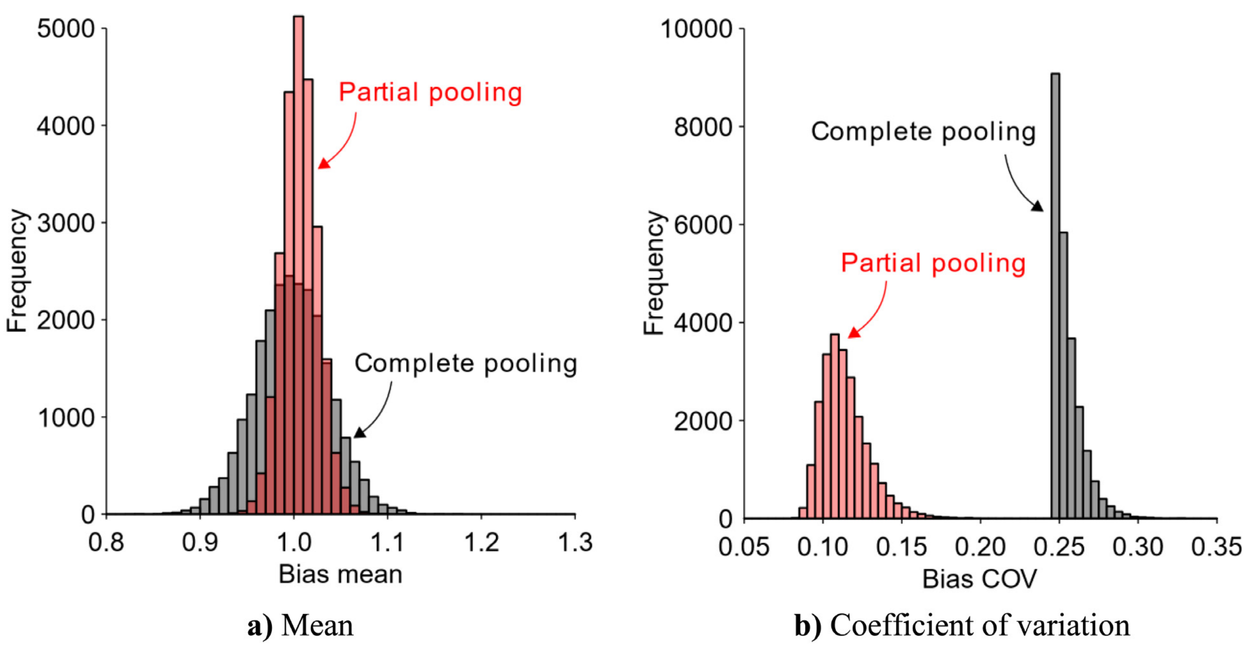

Fig. 6. Posterior bias statistics.

4.1. Predictions

A model fitted in the Bayesian framework using MCMC can be conveniently used to predict pile load capacity. With MCMC simulations available for model parameters, a q_t profile, and with given pile dimensions required as input to Eq. (4a) axial pile capacity can be simulated. Note that the MCMC samples can be from a calibrated model based on generic data, or from a generic model that is updated with site-specific data.

In case of the partial pooling model, at MCMC iteration m, A of a new site A^m is simulated from the population distribution of As by using the sampled μ_A^m and σ_A^m to draw a random value from Lognormal(μ_A, σ_A) (Eq. (4c)). The sampled A^m is used in conjunction with the sampled c^m in Eq. (4a) to calculate mean skin friction for given q_t at each depth. The mean total pile axial capacity due to skin friction is the sum over the depth of the calculated mean skin frictions. The simulated mean total axial capacity profile is further perturbed by error e^m simulated from Eq. (4b) using σ^m. It should be noted that the pile capacity model is calibrated based on total skin friction observations and therefore the residual error is that of the total capacity. Finally, if the complete pooling model is to be used, then there is no need for the two-stage sampling of A^m because there is no "population" of As; samples from posterior A are used directly.

4.2. Updating

As mentioned earlier, the posterior distributions of the parameters of a fitted model can be used as prior distributions for analysis of future data. Bayesian models are dominantly fitted using MCMC, which means usually only simulations from the posterior distributions are available which are not a suitable format for expressing prior distributions. Often, it is possible to obtain reasonable approximations to the MCMC posterior simulations in the form of parametric distributions. This can be achieved by fitting parametric distributions directly to the MCMC

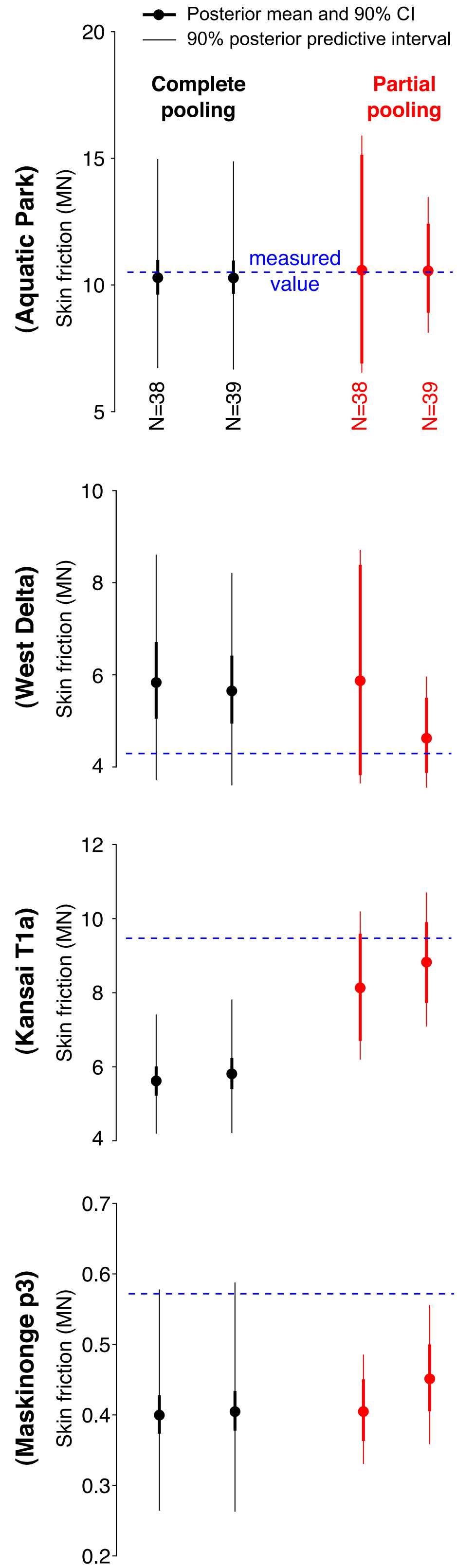

Fig. 7. Leave-one-out predictions from complete and partial pooling models.

simulations or selecting distributions that reasonably capture posterior summary statistics.

For the complete pooling model fitted to all $N = 39$ data, the joint posterior of $\mathrm{Ln}(A)$ and c can be approximated by, for instance, a bivariate normal distribution with mean vector $(-2.88, 0.19)$ and variance–covariance matrix $\begin{pmatrix} 0.035 & 0.012 \\ 0.012 & 0.004 \end{pmatrix}$. The standard deviation σ can be represented as a lognormal distribution with parameters $\mu_\sigma = -1.44$ and $\sigma_\sigma = 0.12$. So, a model with informative priors for analysis of future data could be formulated as:

$$\mathrm{Ln}(\tau_{f_{[i]}}) = \mathrm{Ln}(A) + \mathrm{Ln}(q_{t_{[i]}}) - c \cdot \mathrm{Ln}\left(\left[\max\left(\frac{h_{[i]}}{D^*_{[i]}}, 1\right)\right]\right) + e_{[i]} \tag{5a}$$

$$e_{[i]} \sim \mathrm{Normal}(0, \sigma) \tag{5b}$$

$$\begin{pmatrix} \mathrm{Ln}(A) \\ c \end{pmatrix} \sim \mathrm{MVN}\left(\begin{pmatrix} -2.88 \\ 0.19 \end{pmatrix}, \begin{pmatrix} 0.035 & 0.012 \\ 0.012 & 0.004 \end{pmatrix}\right) \tag{5c}$$

$$\sigma \sim \mathrm{Lognormal}(-1.44, 0.12) \tag{5d}$$

where MVN denotes the multivariate normal distribution.

Prior distributions could also be specified from the partial pooling model. The key difference from the complete pooling model is that parameter A of interest is that of a new site with no data (case “New” in Fig. 2); the resulting prior is weaker (more wide-spread) than that obtained from the complete pooling model because it accounts for between-site variations:

$$\mathrm{Ln}(\tau_{f_{[i]}}) = \mathrm{Ln}(A) + \mathrm{Ln}(q_{t_{[i]}}) - c \cdot \mathrm{Ln}\left(\left[\max\left(\frac{h_{[i]}}{D^*_{[i]}}, 1\right)\right]\right) + e_{[i]} \tag{6a}$$

$$e_{[i]} \sim \mathrm{Normal}(0, \sigma) \tag{6b}$$

$$\begin{pmatrix} \mathrm{Ln}(A) \\ c \end{pmatrix} \sim \mathrm{MVN}\left(\begin{pmatrix} -2.84 \\ 0.19 \end{pmatrix}, \begin{pmatrix} 0.074 & 0.007 \\ 0.007 & 0.003 \end{pmatrix}\right) \tag{6c}$$

$$\sigma \sim \mathrm{Lognormal}(-2.16, 0.17) \tag{6d}$$

Specifying a model similar to Eq. (6) facilitates augmenting of (usually limited) new data (CPT and pile load) with the generic data base via the use of informative priors. However, if it is also desired for the new data to contribute to estimating the hyper-parameters in Eq. (3), i.e. update the prior model of Eq. (6) with the new data for future use, it is more straightforward to include the new data in the generic data base and fit the model from scratch rather than trying to specify a hierarchical model with informative priors obtained from the generic database only. This is generally because of the more complex structure of the posterior distributions of the partial pooling model; a population of As should be specified, mean of which (i.e. μ_A) is correlated with the parameter c.

4.3. Examples

Fig. 7 depicts four selected examples of updating the predictions when site-specific pile load tests are available. The first case presented is from a site with a single pile load test. This is also a pile for which the calculated values (i.e. posterior means) from both models fall close to the measured value. The 90% posterior predictive intervals from the complete and partial pooling models ($N = 38$ analysis) are roughly the same. However, the 90% posterior mean CI from partial pooling is much wider. This is mainly due to the relatively more uncertain coefficient A from the partial pooling model. In the context of the earlier discussion of aleatory and epistemic uncertainties, the complete pooling model shows small epistemic uncertainty (narrow posterior mean) and large aleatory uncertainty. The hierarchical model exhibits a wide posterior CI, indicating large uncertainty about *average* skin friction of a pile in a new site from which we have no observations.

Observing the site-specific data point (i.e. adding the left-out skin friction measurement to the analysis) does not affect predictions from the complete pooling model. This is generally expected (and is also the case for the next three examples) because this model estimates coefficient c and a single coefficient A from all the data ($N = 38$ data points); adding a single observation does not have a considerable effect on the posterior distributions of the parameters. In other words, adding a singe data point does not affect our epistemic uncertainty stemming from the complete pooling model. On the other hand, having a single site-specific observation results in substantial reduction of posterior uncertainty in the partial pooling model. This is essentially the A of a site with no data (New site in Fig. 2) being updated to the A of the Aquatic Park site.

The second case is also from a site with a single observation and shows a similar pattern to the first example. Observing a (smaller than overall average) site-specific data point updates the partial pooling model so that not only posterior uncertainty is reduced, but also that posterior mean is pulled towards the site-specific observation, ruling out larger predicted values for skin friction. As to the first example, the complete pooling model exhibits only minor adjustments in light of the newly observed data point.

The third example in Fig. 7 is one of the outliers detected in Fig. 3. This is a site with two observations. The leave-one-out prediction from the complete pooling model ($N = 38$) fails to capture this observation. Adding back the left-out observation results in a slight increase in the variance of the prediction but not enough to capture this data point. Prior predictions from the partial pooling model ($N = 38$) on the other hand are informed by the second observation from this site, and thus perform better when predicting the left-out data; the updated partial pooling prediction generally shifts towards larger values of skin friction and shows reduced posterior uncertainty.

So far, we discussed examples where the partial pooling model outperformed the complete pooling model. The last example is one where the partial pooling model fails to predict the left-out observation: this is observation 36 in Fig. 4 discussed in Section 3. The left-out data point is from a site (Maskinongé) with three pile load measurements and is included within the 90% predictive interval of the complete pooling prediction (using $N = 38$). On the other hand, the posterior predictive distribution of the partial pooling model severely under-predicts this data point. Including the left-out data point in the analysis results in a slight improvement in the complete pooling prediction. The posterior predictive distribution of the partial pooling model moves towards the site-specific observation and indicates a probability of about 0.03 for this data point. More insight into these predictions can be gained by examining the other two observations from this site (observations 37 and 38 in Fig. 5). These observations exhibit bias values close to unity. On the other hand, the bias for observation 36 is much larger (about 1.4). Such a difference between the biases for data in one site is alarming and could be indicative of issues like site heterogeneity (i.e. the three observations should have not been grouped together for the partial pooling analysis) or perhaps unreliable measurements. Therefore, while it fails to predict the left-out data, the partial pooling model is still useful in that it points out discrepancies that warrant further investigation. Pursuing such an investigation is beyond the scope of this paper. Nevertheless, it is worth mentioning that the Maskinongé pile load tests are amongst those with the smallest weights (i.e. subjective measure of load test reliability) assigned by the JIP Team of experts assessing the quality of the pile load tests included in the Unified database [19].

The above leave-one-out analyses are examples of one of the features that makes the Bayesian framework attractive for geotechnical probabilistic design, namely, Bayesian updating. Bayesian updating has been discussed in the geotechnical literature, mostly with focus on the underlying theory and concepts and Bayesian computation (e.g. [12,33,34]). On the other hand, systematic and formal formulation of geotechnical prior distributions is less-studied, and priors are constructed (often simplistically) from historical parameter ranges and summary statistics. Complete and no pooling models are also sometimes used (e.g. [34]).

The above analyses illustrate the importance of quantifying the uncertainties in the context of updating a model with site-specific data. The analyses also emphasise how the allocation of uncertainties in the two complete and partial pooling models results in different prior distributions and subsequently different (updated) posteriors. They particularly highlight how the assumption of zero between-site variation in complete pooling models results in narrow priors that are good overall representations of the analysed database but hardly exploit newly observed data. The hierarchical model on the other hand explicitly models between-site variations, resulting in wider – but still informative – prior distributions that can be meaningfully updated in view of observing data from a new site.

5. Summary and conclusions

This paper explored formulating geotechnical model calibration as a hierarchical Bayesian regression problem. A model and database for axial capacity of piles in clay were used as an example. The analyses focused on two aspects of uncertainty quantification that are often overlooked in the geotechnical engineering literature, namely, statistical parameter uncertainty and potential variations between sub-groups of a generic database.

The fitted hierarchical model was compared with the more commonly used complete pooling model, and it was shown to provide a more nuanced assessment and quantification of uncertainties by allowing group-specific parameters that vary from one another while simultaneously borrow information from each other. Furthermore, examples from a leave-one-out analysis revealed limitations of the complete pooling modelling approach particularly when it comes to combining past experience (i.e. a generic database) with (potentially limited) project or site-specific data. The hierarchical model on the other hand was shown to be better suited for summarising the generic database in the form of prior distributions that do not over-fit the database and thus can be meaningfully updated considering limited project or site-specific data.

Although available for decades, hierarchical Bayesian modelling has only recently gained some popularity in the field of probabilistic geotechnical engineering. Results from this paper and other recent work (e.g. [8,9,17,18]) highlight the relevance and potential of hierarchical modelling for becoming a standard approach for analysing geotechnical engineering data, particularly in the context of the ongoing challenge of formulating informative prior distributions from past data and experience. We believe this potential warrants establishing protocols and guidelines tailored to analysis of geotechnical data, and initiatives for re-analysing available data using hierarchical models.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The work described in this paper was done while the first author was a post-doctoral fellow at NGI granted by the Natural Sciences and Engineering Council of Canada (NSERC) (grant PDF-532979-2019). Their support is gratefully acknowledged. Part of the work in this paper was done in the JIP “Reliability of axial pile capacity calculation methods” carried out by NGI and the University of Western Australia, UWA. The JIP was sponsored by Equinor AS, Lundin Norway AS, Ørsted, DNV AS, ONGC and Petrobras. The work was continued in the JIP “Unified, ‘non-affiliated’ CPT-based axial pile capacity design method”, with the additional sponsors: BP, Total, ExxonMobil, EnBW, EDF, HDEC and Aramco. The support of the JIP sponsors is gratefully acknowledged. The opinions expressed in this paper are those of the authors and do not necessarily represent the viewpoints of the JIP sponsors.

Table A.1
Details of pile load test in Unified database (from [22]).

N	Site name	Test	Type	Borehole depth (m)	Tip depth (m)	t_{eq} (days)	D (m)	D_i (m)	L/D	Q_m (MN)
1	Onsoy	A1–02	CET	5.0	15	26	0.22		45.7	0.091
2	Onsoy	A3–02	CET	20.0	30	54	0.22		45.7	0.224
3	Onsoy	B1–02	OET	5.0	15	81	0.81	0.79	12.3	0.427
4	Onsoy	C1–02	CET	5.0	35	50	0.22		137.0	0.407
5	Onsoy	C2–02	CET	5.0	35	51	0.22		137.0	0.487
6	Pentre	A6-02a	CET	25.0	32.5	32	0.22		34.2	0.351
7	Pentre	LDP	OEC	15.0	55	44	0.76	0.73	52.5	6.32
8	Tilbrook	A1	CET	3.0	12.9	61	0.22		45.2	1.246
9	Tilbrook	B1	CET	17.5	25.6	59	0.22		37.0	1.741
10	Tilbrook	C1	CET	3.0	17.5	59	0.22		66.2	2.045
11	Tilbrook	D1	OET	3.0	17.5	73	0.27	0.24	53.1	2.039
12	Tilbrook	LDP-C	OEC	0.0	30	130	0.76	0.70	39.4	16.5
13	Cowden	A	OEC	0.0	9.2	30	0.46	0.42	20.1	1.18
14	Cowden	B	CEC	0.0	9.2	30	0.46		20.1	1.42
15	AquaticPark	S2–1	OET	57.9	80.5	60	0.76	0.69	29.7	10.5
16	Kontich	B	OEC	1.5	23.5	21	0.61	0.56	36.1	5.07
17	Kansai	T1a	OEC	0.0	32.8	35	1.50	1.46	21.9	9.47
18	Kansai	T2	OEC	0.0	48.3	42	1.50	1.46	32.2	17.00
19	SintKatelijne	A1	CEC	1.0	7.4	92	0.35		18.3	0.975
20	SintKatelijne	A4	CEC	1.0	11.6	89	0.35		30.3	1.66
21	WestDelta	LS1	OET	0.0	71.3	116	0.76	0.72	93.6	4.29
22	Onsoy2	O1–1	OET	1.4	19.1	78	0.51	0.50	34.8	0.519
23	Cowden2	C2–1	OET	1.0	10	119	0.46	0.43	19.7	1.02
24	Femern	F2–1	OET	0.0	25	34	0.51	0.47	49.2	3.12
25	Stjordal	S2–1	OET	1.0	23.6	50	0.51	0.50	44.5	0.64
26	Merville	D1	OEC	0.0	9.4	44	0.51	0.48	18.5	1.165
27	Merville2	B1S1	CEC	4.0	13	57	0.41		22.2	1.55
28	Merville2	B3S1	CET	4.0	13	62	0.41		22.2	1.4
29	Klang	TP1A	OEC	0.0	35.5	26	0.25	0.14	142.0	0.635
30	Riau	G1-T1	OEC	0.0	24	73	0.35	0.20	68.6	0.425
31	Riau	G10-T1	OEC	0.0	30	71	0.35	0.20	85.7	0.5
32	Riau	G6-T1	OEC	0.0	36	68	0.35	0.20	102.9	0.7
33	GoldenEars	SC	CEC	0.0	36	120	0.36		100.8	2.8
34	LuluIsland	UBC1	CEC	2.0	14.3	82	0.32		38.0	0.225
35	Quebec	9	CET	2.3	18.1	66	0.32		49.5	0.426
36	Maskinongé	p3	CEC	0.0	23.8	58	0.23		103.5	0.61
37	Maskinongé	p4	CEC	0.0	23.8	58	0.22	0.21	108.7	0.4
38	Maskinongé	p5	CEC	0.0	37.5	58	0.23		163.0	0.88
39	Goteborg	a	CEC	2.0	18	34	0.24		68.1	0.23

Appendix. Pile load test database

See Table A.1.

References

[1] W.S. Sillers, D.G. Fredlund, Statistical assessment of soil–water characteristic curve models for geotechnical engineering, Can. Geotech. J. 38 (6) (2001) 1297–1313.

[2] T.M. Allen, R.J. Bathurst, Improved simplified method for prediction of loads in reinforced soil walls, J. Geotech. Geoenviron. Eng. 141 (11) (2015) 04015049.

[3] M. D'Ignazio, K.K. Phoon, S.A. Tan, T.T. Länsivaara, Correlations for undrained shear strength of finnish soft clays, Can. Geotech. J. (53) (2016) 1628–1645.

[4] P. Paniagua, M. D'Ignazio, J.S. L'Heureux, T. Lunne, K. Karlsrud, CPTU correlations for norwegian clays: An update, AIMS Geosci. 5 (2) (2019) 82–103, http://dx.doi.org/10.3934/geosci.2019.2.82.

[5] F.M. Naftchali, R.J. Bathurst, Influence of geosynthetic stiffness on analytical solutions for reinforced fill over void, Geosynth. Int. (Online) (2021) http://dx.doi.org/10.1680/jgein.21.00015a.

[6] J.G. Surles, H. Seo, J. Doerzbacher, W.D. Lawson, P.W. Jayawickrama, Confidence intervals for resistance factors in geotechnical LRFD applications, Struct. Saf. 80 (2019) 1–12.

[7] N. Bozorgzadeh, R.J. Bathurst, A Bayesian approach to reliability-based analysis of MSE walls considering parameter uncertainty, Georisk 15 (1) (2021) 1–15.

[8] N. Bozorgzadeh, R.J. Bathurst, Hierarchical Bayesian approaches to statistical modelling of geotechnical data, Georisk (Online) (2020) http://dx.doi.org/10.1080/17499518.2020.1864411.

[9] J. Ching, S. Wu, K.K. Phoon, Constructing quasi-site-specific multivariate probability distribution using hierarchical Bayesian model, J. Eng. Mech. 147 (10) (2021) 04021069.

[10] A. Gelman, H.S. Stern, J.B. Carlin, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, CRC Press, Boca Raton, FL, USA, 2013.

[11] D. Lunn, C. Jackson, N. Best, D. Spiegelhalter, A. Thomas, The BUGS Book: A Practical Introduction to Bayesian Analysis, CRC Press, Boca Raton, FL, USA, 2012.

[12] J. Zhang, P. Boothroyd, M. Calvello, M. Eddleston, A.C. Grimal, P. Iason, Z. Lou, S. Najjar, A. Rodriguez-Marek, D. Straub, M. Uzielli, Y. Wang, Bayesian Method: A Natural Tool for Processing Geotechnical Information, Joint TC205/TC304 Working Group on Discussion of Statistical/Reliability Methods for Eurocodes Rep., International Society for Soil Mechanics and Geotechnical Engineering, London, UK, 2017.

[13] G.B. Baecher, Bayesian thinking in geotechnics, in: D.V. Griffiths, G.A. Fenton, J. Huang, L. Zhang (Eds.), Geo-Risk 2017: Keynote Lectures, in: Geotechnical Special Publication, vol. 282, American Society of Civil Engineers, Reston, VA, USA, 2017, pp. 1–18.

[14] J. Zhang, J.P. Li, L.M. Zhang, H.W. Huang, Calibrating cross-site variability for reliability-based design of pile foundations, Comput. Geotech. 62 (2014) 154–163.

[15] J. Zhang, H. C. Juang, Martin. J.R., H.W. Huang, Inter-region variability of robertson and wride method for liquefaction hazard analysis, Eng. Geol. 203 (2016) 191–203.

[16] N. Bozorgzadeh, J.P. Harrison, M.D. Escobar, Hierarchical Bayesian modelling of geotechnical data: Application to rock strength, Geotechnique 69 (12) (2019) 1056–1070.

[17] J. Ching, K.K. Phoon, Constructing site-specific multivariate probability distribution model using Bayesian machine learning, J. Eng. Mech. 145 (1) (2019) 04018126.

[18] J. Ching, K.K. Phoon, Z. Yang, A.W. Stuedlein, Quasi-site-specific multivariate probability distribution model for sparse, incomplete, and three-dimensional spatially varying soil data, Georisk Assess. Manag. Risk Eng. Syst. Geohazards (2021) 1–24.

[19] B. Lehane, Z. Liu, E. Bittar, F. Nadim, S. Lacasse, R. Jardine, P. Carotenuto, M. Rattley, P. Jeanjean, K. Gavin, R. Gilbert, J. Bergan-Haavik, N. Morgan, A new 'unified' CPT-based axial pile capacity design method for driven piles in sand, in: Z. Westgate (Ed.), 4th International Symposium on Frontiers in Offshore Geotechnics, Austin, Texas. 2022, Deep Foundations Institute, Hawthorne, 2020.

[20] B.M. Lehane, J.K. Kim, P. Carotenuto, F. Nadim, S. Lacasse, R.J. Jardine, B.F.J. van Dijk, Characteristics of unified databases for driven piles, in: Offshore Site Investigation and Geotechnics, Smarter Solutions for Future Offshore Developments: Proceedings of the 8th International Conference, 12-14 2017, Royal Geographical Society, London, UK, 2017, pp. 162–191, Society for Underwater Technology.

[21] Z. Liu, F. Nadim, S. Lacasse, B. Lehane, Y.J. Choi, Method uncertainty for five axial pile capacity design methods, in: Offshore Technology Conference, OTC 2019, 6-9 2019, Houston, #OTC-29514-MS, 2019.

[22] Norwegian Geotechnical Institute, A New CPT-Based Design Method for Assessment of Axial Capacity of Driven Piles in Clay, NGI Report 20180089-02-R, 2021.

[23] B.M. Lehane, Z. Liu, E.J. Bittar, F. Nadim, S. Lacasse, N. Bozorgzadeh, et al., CPT-based axial capacity design method for driven piles in clay, J. Geotech. Geoenviron. Eng. 148 (9) (2022).

[24] P.K. Robertson, Interpretation of cone penetration tests—A unified approach, Can. Geotech. J. 46 (11) (2009) 1337–1355.

[25] A. Gelman, D. Simpson, M. Betancourt, The prior can often only be understood in the context of the likelihood, Entropy 19 (10) (2017) 555.

[26] Stan Development Team, RStan: The R interface to Stan, 2019, R package version 2.21.0. (http://mc-stan.org).

[27] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021.

[28] N. Bozorgzadeh, R.J. Bathurst, Bayesian model checking, comparison and selection with emphasis on outlier detection for geotechnical reliability-based design, Comput. Geotech. 116 (2019) 103181.

[29] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models, Stat. Comput. 24 (6) (2014) 997–1016.

[30] J. Ching, G.-H. Lin, J.-R. Chen, K.-K. Phoon, Transformation models for effective friction angle and relative density calibrated based on generic database of coarse-grained soils, Can. Geotech. J. 54 (4) (2017) 481–501, http://dx.doi.org/10.1139/cgj-2016-0318.

[31] A. O'Hagan, J.E. Oakley, Probability is perfect, but we can't elicit it perfectly, Reliab. Eng. Syst. Saf. 85 (1–3) (2004) 239–248.

[32] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2) (2009) 105–112.

[33] D. Straub, I. Papaioannou, Bayesian analysis for learning and updating geotechnical parameters and models with measurements, in: K.K. Phoon, J. Ching (Eds.), Risk and Reliability in Geotechnical Engineering, ch. 5, CRC Press, Boca Raton, FL, USA, 2014, pp. 221–264.

[34] Z.J. Cao, Y. Wang, D.Q. Li, Site-specific characterization of soil properties using multiple measurements from different test procedures at different locations—A Bayesian sequential updating approach, Eng. Geol. 211 (2016) 150–161.