Hierarchical Bayesian modelling for geotechnical parameter derivation

Mavritsakis, A.; Schweckendiek, T.; Teixeira, A.M.; Smyrniou, E.

DOI
10.3850/978-981-18-5184-1_MS-13-037-cd

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management

Proc. of the 8th Intl. Symp. on Reliability Engineering and Risk Management (ISRERM 2022)
Edited by Michael Beer, Enrico Zio, Kok-Kwang Phoon, and Bilal M. Ayyub
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5184-1_MS-13-037-cd

Hierarchical Bayesian modelling for geotechnical parameter derivation

A. Mavritsakis*[1], T. Schweckendiek[1,2], A. M. Teixeira[1], and E. Smyrniou[1]

[1]*Deltares, Delft, The Netherlands*
[2]*Department of Hydraulic Delft University of Technology, The Netherlands*

Abstract

Bayesian inference poses as a means for characterizing the uncertainty in geotechnical parameters based on limited site investigation data. In this study, a Hierarchical Bayesian analysis framework is used to analyse a site investigation database in order to derive geotechnical soil parameters for two widely applied strength models. The first one focuses on calibrating the relationship between in-situ CPT measurements and undrained shear strength. The second one is the SHANSEP soil strength model, which is used for evaluating the undrained shear strength using OCR information. The framework operates in a hierarchical fashion, performing inference on separate project sites and at the same time drawing conclusions on a global level. The result is site characterization on a probabilistic level and the derivation of geotechnical parameters together with their probability distributions. The results are assessed by evaluating their influence in the failure probability of a geotechnical structure, demonstrating that the proposed hierarchical approach provides a more complete description of uncertainty than standard practice methods.

Keywords: Hierarchical Bayesian Modelling, Bayesian Inference, SHANSEP, Site Characterisation, Soil strength, Uncertainties

1. Introduction

While laboratory testing can provide direct estimates of geotechnical parameters for the design or assessment of structures, in-situ tests are performed on a greater scale and draw information on the undisturbed behaviour and the spatial heterogeneity of the soil. In these cases, Bayesian inference can pose as a means for calibrating the relationships between in-situ and laboratory test data, as well as for characterizing the uncertainty and spatial variability of soil parameters. In this study, a Hierarchical Bayesian modelling (HBM) is explored to establish the relationship between the CPT (Cone Penetration Test) in-situ measurements and the undrained shear strength of the soil as determined in the lab. The same approach is used to calibrate the SHANSEP strength parameters, which are used for modelling the undrained shear strength of certain clay soils, and to identify the spatial heterogeneity pattern of soil parameters.

In this context, the impact of HBM is examined on two different geotechnical parameter settings within the undrained shear strength parameter estimation. These HBM applications are showcased and compared to a standard practice approach.

The HBM operates in a hierarchical fashion, performing inference on separate project sites and at the same time drawing conclusions for the parameter population. The result is site characterization on a probabilistic level and the derivation of geotechnical parameters together with their probability distributions. This outcome can be utilized in determining characteristic values, or in a fully probabilistic analysis in a project site and on a regional level.

Section 1 and 2 are introductory sections to the problem and the theory behind HBM, respectively. Section 3 presents the benefits of adopting HBM in deriving geotechnical parameters by analysing the structure of the statistical models and comparing the resulting posteriors. Section 4 investigates the advantages of employing HBM by evaluating its impact on the assessment of probability of failure for an existing embankment. Finally, Section 5 presents the conclusions and follow ups.

2. Theoretical background and approach

2.1 Bayesian inference

Bayesian inference is used to draw conclusions on variables via data and observations by employing Bayes' theorem (Eq. 1). It entails the set-up of a statistical model, which describes initial information on the variables through the prior distribution. A key component of the model is the formulation of the likelihood function, which measures the competency of the model in describing the observations. Conditioning on the observations leads to a statistical model that expresses the updated knowledge on the variables via the posterior distribution. Note that Bayesian inference reduces epistemic uncertainty in the statistical model, while aleatory uncertainty remains.

$$P(\theta|\varepsilon) = \frac{L(\varepsilon|\theta)*P(\theta)}{\int P(\varepsilon|\theta)*P(\theta)d\theta} \quad (1)$$

In the above equation:

- θ is the variable vector that is being inferred;
- ε is the vector of observations;
- $P(\theta|\varepsilon)$ is the posterior distribution;
- $P(\theta)$ is the prior distribution;
- $L(\varepsilon|\theta)$ is the likelihood of the observations given the variable vector.

In the analyses presented in this paper, Bayesian inference is calculated using the *No U-Turn Sampler* (NUTS) algorithm [1], which is a highly efficient version of Hamiltonian Monte Carlo (HMC). The algorithm is implemented via the Probabilistic Programming Python package PyMC3 [2].

2.2 Robust Bayesian Linear Regression

Bayesian inference can be employed to fit Linear Regression models. A typical problem experienced in such models in the sensitivity of the fit to data outliers. To mitigate this impact, Robust Linear Regression techniques are adopted [3]. In the context of this paper, Robust Linear Regression models are implemented by assuming that the regression errors follow t-Student distributions, instead of Normally distributed ones. The t-Student distribution is able to allocate higher likelihood values at far-off points,

* E-Mail: Antonis.Mavritsakis@deltares.nl

interpreting outliers as high-variance observations and reducing their impact on the regression parameters' posteriors. In addition to the mean (μ) and standard deviation (σ), the t-Student distribution is parametrized by the number of degrees of freedom (ν_{dof}), which controls the behaviour of the distribution's tails.

2.3 Hierarchical Bayesian Modelling

In several engineering problems, variables that describe the same phenomenon at different groups of data are implicitly related, as compelled by the data generating process. Consequently, the variables of each group belong to the same population. Bayesian models can be divided in three types when it comes to the treatment of grouped data:

- Pooled models treat all data as one group or population, without any distinction between groups.
- Unpooled models make distinctions between groups, but groups are independent.
- Hierarchical models (partially pooled) make distinctions between groups and a level of dependency among groups is included.

A benefit of adopting Bayesian inference is the ability to operate in hierarchical structures. Hierarchical Bayesian Models (HBM) infer the variables per group, as well as the parameters defining the population distributions through hyperparameters. The hyperparameters (φ) describe the population distribution of the group variables (θ) and infer them by conditioning the HBM on the observations (ε) as indicated in Eq. 2. The simplification holds, because the observations depend directly only on the group variables, while the two-way influence between the observations and the hyperparameters is indirect and mediated by the group variables [3]. Additionally, the group variables interact via a common dependency on the hyperparameters, which means that observations for group ε_i provide information for all group variables θ, instead of only affecting the variables of their group θ_i during inference.

$$P(\varphi,\theta|\varepsilon) = \frac{P(\varepsilon|\theta)*P(\varphi,\theta)}{\int P(\varepsilon|\theta)*P(\varphi,\theta)d\varphi d\theta} \quad (2)$$

A noteworthy feature of HBMs for geotechnical engineering is their ability to infer the population posterior distribution. Considering that geotechnical data is usually grouped per site, the posterior distribution of the population can express information for site variables on a regional, national, or even global level, depending on the origin of the data. The population posterior distributions are expected to enhance predictions at a new site, where no observations are yet available.

HBM is flexible enough to incorporate the Robust Linear Regression scheme. This combination, which belongs to the family of Hierarchical Linear Models, will be used in the following sections of this paper.

Fig. 1 provides a generic description of the adopted HBM structure. A set of three hyperparameters influences the regression coefficients per group (a_i and β_i). These are the mean (μ) and standard deviation (σ) per regression coefficient, as well as their correlation coefficient (ρ). The regression coefficients per group are used to evaluate the regression model of the group, which is then applied in the likelihood function. Two more hyperparameters, the standard error of the regression (σ_ε) and the number of degrees of freedom (ν) - which is required in the Robust Linear Regression scheme - are used to evaluate the likelihood function of the group. Since groups are assumed to be independent, the likelihood of the model is the product of the individual group likelihoods (Eq. 3)

$$L_{model}(\varepsilon) = \prod_{i=1}^{n_{groups}} L(\varepsilon_i|\theta_i) \quad (3)$$

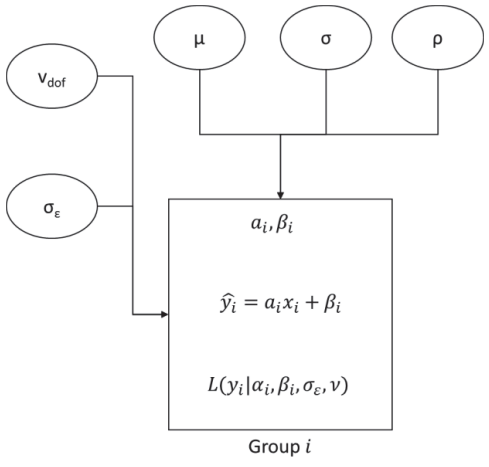

Figure 1. Illustration of the general HBM structure used in the study.

2.4 Geotechnical background

The impact of HBM is examined on two different geotechnical parameter derivation settings, which are organics parts of standard geotechnical practice. Hence, they pose as candidates for showcasing the impact of adopting HBM over the standard practice approach, as well as the pooled and unpooled Bayesian models. Specifically, the settings are:

- The N_{kt} regression, which connects the corrected CPT cone resistance (q_{net}) to the undrained shear strength of clays (S_u), according to Eq. 4 [4]. This relationship has high practical value, as it allows for mapping S_u over the subsoil by utilizing the abundance of CPT measurements in site investigations datasets.

$$q_{net} = N_{kt} * S_u \quad (4)$$

- The SHANSEP strength model regression, which is defined by parameters S and m [4]. SHANSEP connects the maximum effective stress level and the over-consolidation ratio (OCR) to the undrained shear strength of clays (S_u), according to Eq. 5 Consequently, Eq. 6 is derived and reformulates the SHANSEP strength model parameters as the coefficients of a linear regression model; log(S) takes the role of the intercept while m is the slope coefficient.

$$S_u = S * \sigma' * OCR^m \quad (5)$$

$$\log\left(\frac{S_u}{\sigma'}\right) = \log(S) + m * \log(OCR) \quad (6)$$

The SHANSEP strength model is widely used in geotechnical calculations of clay and peat soils, so an improvement of the derivation of its parameters can prove valuable.

2.5 Inference procedure set-up

The world-wide *CLAY10* database [5], containing data from CPT measurements and laboratory tests (totalling 374 datapoints gathered at 33 sites over the globe) is used in this study to provide the observations for the inference procedure.

For judging the inference results, the cross-validation method is adopted, which according to [3] entails that the database is in a training dataset for inference and a testing dataset for evaluating the predictive power of the inferred model. In this study, 80% to 20% of the total observations are attributed to the training and testing datasets respectively.

Two metrics are used to evaluate the prowess of the HBM. The first metric is the log-likelihood of the test observations, as calculated with the posterior predictive distribution, as described by Eq. 7.

$$\log L = \sum_{i=1}^{n} \log \int p(\varepsilon_i^{test}|\theta) p(\theta|\varepsilon) d\theta \quad (7)$$

The second metric is the expected value Root Mean Square Error (RMSE), which is defined according to Eq. 8.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i^{test} - \bar{\varepsilon}_i^{test})^2} \quad (8)$$

In this analysis, the NUTS algorithm is executed with 4 chains and 2,500 posterior samples. This adjustment has been examined to sufficiently explore the variable domain.

3. Bayesian estimation of geotechnical parameters

3.1 Standard practice

The standard practice approach in the Netherlands [4] comprises the derivation of the N_{kt} distribution, as the slope coefficient of a regression analysis with zero intercept using the Ordinary Least Squares (OLS) method. Even though the method leads to distributions for the regression coefficient, as well as a description of the regression error, it is still deterministic, because it derives just the point estimates of the distribution parameters.

The standard approach for deriving the SHANSEP parameters consists of a Linear Regression analysis which yields the SHANSEP parameters and their distributions. It should be noted that in most practical applications, the *m* parameter is derived by oedometer tests, and the regression analysis is performed for the S parameter. However, since no oedometer tests are available in the database, the forementioned procedure is adopted.

3.2 HBM inference of the N_{kt} regression model

In the Bayesian models, N_{kt} is determined as the slope coefficient of a Robust Linear Regression model fit between q_{net} and S_u. The regression model is selected to have a zero-intercept coefficient for the sake of compatibility with standard practice. In this way, N_{kt} reflects part of the reducible uncertainty of the model, while the regression error represents irreducible uncertainty. Moreover, the regression follows a heteroscedastic interpretation of the errors, which is supported by the shape of the point clouds and the physical phenomenon modelled; the strength uncertainty is proportional to the cone resistance and when the cone resistance is zero, the soil has no strength. So, the Bayesian model infers the Coefficient of Variation (CoV), which is used to define the error regression model estimate ($\widehat{S_u}$) using Eq. 9. Lastly, since the Robust Linear Regression scheme is adopted, the number of degrees of freedom (ν_{dof}) of the t-Student likelihood function should be inferred.

$$\sigma_{error} = CoV\ \widehat{S_u} \quad (9)$$

Tab. 1 presents the priors for the random variables of the HBM. The priors of the hyperparameters $\mu_{N_{kt}}$ and $\sigma_{N_{kt}}$ are weakly informative. It is a-priori known that N_{kt} takes non-negative values. Thus, it follows a truncated normal distribution. However, variable N_{kt} is indirectly sampled with an auxiliary variable ($N_{kt,offset}$) [6], to allow for better exploration of the variable domain. Moreover, the CoV and ν_{dof} are the same for all sites. This choice might lead to some inaccuracy in modelling aleatory uncertainty per site but improves largely the applicability of the results to new sites.

Table 1. Prior distributions for the random variables of the HBM of the Nkt setting.

Variable	Prior	
	Type	Parameters
$\mu_{N_{kt}}$	Half normal	$\mu = 0$ $\sigma = 10$
$\sigma_{N_{kt}}$	Half normal	$\mu = 0$ $\sigma = 10$
$N_{kt,offset}$	Truncated normal	$\mu = 0$ $\sigma = 1$ $lower = -\frac{\mu_{N_{kt}}}{\sigma_{N_{kt}}}$
CoV	Half normal	$\mu = 0$ $\sigma = 1$
ν_{dof}	Inverse gamma	$\alpha = 1$ $\beta = 1$

$$N_{kt} = \mu_{N_{kt}} + \sigma_{N_{kt}}\ N_{kt_{offset}} \quad (10)$$

$$\widehat{S_u} = q_{net}/N_{kt} \quad (11)$$

$$L(S_u|N_{kt}, CoV_{error}) = \prod_{i=1}^{n_{groups}} \prod_{j=1}^{n_i} t - Student(S_{u_{i,j}}|\mu = \widehat{S_{u_{i,j}}}, \sigma = \widehat{S_{u_{i,j}}}\ CoV_{error}, \nu = \nu_{dof}) \quad (12)$$

Eq. 10 – 12 describe the derivation of the regression coefficient N_{kt} from the random variables and the formulation of the regression model and the likelihood function per site. As shown by Eq. 3, the likelihood of the model is the product of the individual site likelihoods. Also, the likelihood per site is the product of the likelihood per data point in the site, as the points are assumed to be

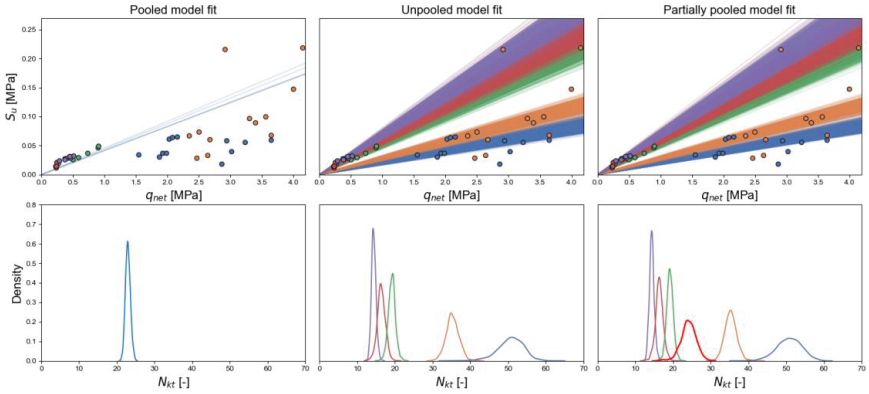

Figure 2. Posterior regression model fits for the N_{kt} model (sites indicated by colour) and KDE plots of the posterior samples for the pooled, unpooled and partially pooled models (only 5 sites displayed for clarity).

independent. Thus, Eq.12 describes the likelihood of the model as this double product, with index i indicating the site and index j indicating the datapoint of each site.

Fig. 2 presents the regression lines and Kernel Density Estimation (KDE) [7] fits of the posterior samples for each of the pooled, unpooled and partially pooled models. The stiff behaviour of the pooled model does not allow the regression line to follow the data, Thus, it leads to low uncertainty of the regression slope, but compensates for the uncertainty of the data through a high standard error, which resembles aleatory uncertainty. The HBM displays similar behaviour to the unpooled model. However, its posterior distributions are just slightly wider, indicating a greater extent of uncertainty in its variable estimations, which is introduced by the incorporation of hyperparameters. Expectedly, the posterior of the population mean hyperparameter of the HBM is centred at the same values as the N_{kt} posterior of the pooled model, since it needs to capture the trend of all sites. Yet, the former has greater spread to allow for site-specific N_{kt} variables to approximate the values that best describe their datasets.

Table 2. Log-likelihood score and expected RMSE per model on the train and test datasets for the N_{kt} regression model.

Model	Log-likelihood [-] on the:		RMSE [MPa] on:	
	Train dataset	Test dataset	Train dataset	Test dataset
Pooled	2.98	-5.26	7.0E-4	5.2E-4
Unpooled	3.81	3.58	2.3E-4	3.2E-4
Partially pooled	3.81	3.58	2.3E-4	3.2E-4

Tab. 2 presents the metric scores of the pooled, unpooled and partially pooled Bayesian models. It should be noted that the log-likelihood metric is scaled by the number of entries in the dataset. In this way, the loglikelihood scores of the train and test datasets are comparable. Evidently, the pooled model is the least flexible and achieves the worst scores in terms of both RMSE and log-likelihood. On the other hand, the unpooled and HBM models score equally well in the RMSE and log-likelihood metrics for the training and testing dataset. In this case, the HBM closely approximates the unpooled model. Although the two models have the same effectiveness, the unpooled model lacks the derivation of population-level conclusions.

3.4 HBM inference of the SHANSEP regression model

Tab. 3 presents the prior distributions attributed to each of the random variables of the HBM for the SHANSEP setting. The model is defined similarly to the respective model of the previous section, but in this case both regression coefficients are being inferred. Again, weakly informative priors are selected for the hyperparameters. Since this regression model includes two coefficients, the correlation between them needs to be inferred too. Otherwise, the model is incapable of dependably sampling regression coefficients at a population level. The correlation matrix has been attributed a weakly informative prior, modelled with the Lewandowski-Kurowicka-Joe distribution (LKJ) [8]. The regression slope coefficient offset are sampled from a multivariate distribution that is truncated, in order to ensure that m is non-negative. As in the previous section, the variable that models irreducible uncertainty, the standard error (σ_{error}), as well as the number of degrees of freedom of the likelihood function (ν_{dof}), are common for all sites in the partially pooled model, in order to increases the utility of the results in making predictions for new sites.

Eq. 13-14 describe the derivation of the regression coefficients α and β. from the random variables. Eq. 15-16 formulate the regression model and the likelihood function of the HBM, which is expressed similarly to the one for the N_{kt} regression setting.

$$\alpha = \mu_\alpha + \sigma_\alpha \, \alpha_{offset} \quad (13)$$

$$\beta = \mu_\beta + \sigma_\beta \, \beta_{offset} \quad (14)$$

$$\hat{y} = \alpha + \beta \, x \quad (15)$$

$$L(y|\alpha,\beta,\sigma) = \prod_{i=1}^{n_{sites}} \prod_{j=1}^{n_i} t - Student\left(y_{i,j} \middle| \mu = \hat{y}_{i,j}, \sigma = \sigma_{error}, \nu = \nu_{dof}\right) \quad (16)$$

Table 3. Prior distributions for the random variables of the HBM of the SHANSEP setting.

Variable	Prior	
	Type	Parameters
μ_a	Normal	$\mu = 0$ $\sigma = 5$
σ_a	Half normal	$\mu = 0$ $\sigma = 5$
μ_β	Half normal	$\mu = 0$ $\sigma = 2$
σ_β	Half normal	$\mu = 0$ $\sigma = 1$
$\Sigma_{\alpha\beta}$	LKJ	$\eta = 2$
$\begin{bmatrix}\alpha_{offset}\\ \beta_{offset}\end{bmatrix}$	Truncated normal	$\mu = 0$ $\Sigma = \Sigma_{\alpha\beta}$ $lower = \begin{bmatrix}-\infty\\ -\frac{\mu_\beta}{\sigma_\beta}\end{bmatrix}$
σ_{error}	Half normal	$\mu = 0$ $\sigma = 1$
ν_{dof}	Inverse gamma	$\alpha = 1$ $\beta = 1$

Fig. 3 presents the regression lines and KDE posterior plots for each of the pooled, unpooled and partially pooled models. The pooled model shows low uncertainty in the regression coefficients and greater values of the error standard deviation. In other words, the pooled model leads to a greater extent of aleatory uncertainty. On the other hand, the unpooled model is ill-behaved, as evidenced by the great spread of the SHANSEP parameter distributions and unstructured behaviour of the regression lines. This stems from the nature of data available per site and manifests as combinations of the regression coefficients that span a wide value ranges being able to describe the data sufficiently well. Examples of the data characteristics that lead to the forementioned behaviour is the lack of over-consolidated samples in a site dataset, which expectedly leads to no information for the regression slope, or OCR values that are confined in narrow ranges. Lastly, the HBM is flexible enough to sufficiently reduce epistemic uncertainty and provide a satisfactory fit per site. Additionally, the hyperparameters impact the prior of the site parameters compel the regression coefficients to follow a global trend. This leads to well behaved posteriors that are concentrated in explicit neighbourhoods of the S-m plane, even with data of unfavourable nature.

Tab. 4 presents the metric scores of the pooled, unpooled and partially pooled Bayesian models. The unpooled model scores significantly better RMSE values in the training dataset than in the testing dataset, a behaviour that usually implies overfitting. On the other hand, both RMSE values are consistent and sufficiently low for the HBM (0.06 for the training dataset and 0.07 for the testing dataset). Also, the HBM achieves a greater loglikelihood score in the testing dataset than the unpooled model, even though the opposite results are met in the training dataset. This suggests that the stronger structure of the partially pooled model leads to better predictions than the unpooled model. Due to its rigidity, the pooled model lacks accuracy and exhibits low predictive power.

Table 4. Log-likelihood score and expected RMSE per model on the train and test datasets for the SHANSEP regression model.

Model	Log-likelihood [-] on the:		RMSE [kPa] on:	
	Train dataset	Test dataset	Train dataset	Test dataset
Pooled	-0.51	-0.60	0.17	0.19
Unpooled	0.25	-0.06	0.07	0.20
Partially pooled	0.20	0.05	0.06	0.07

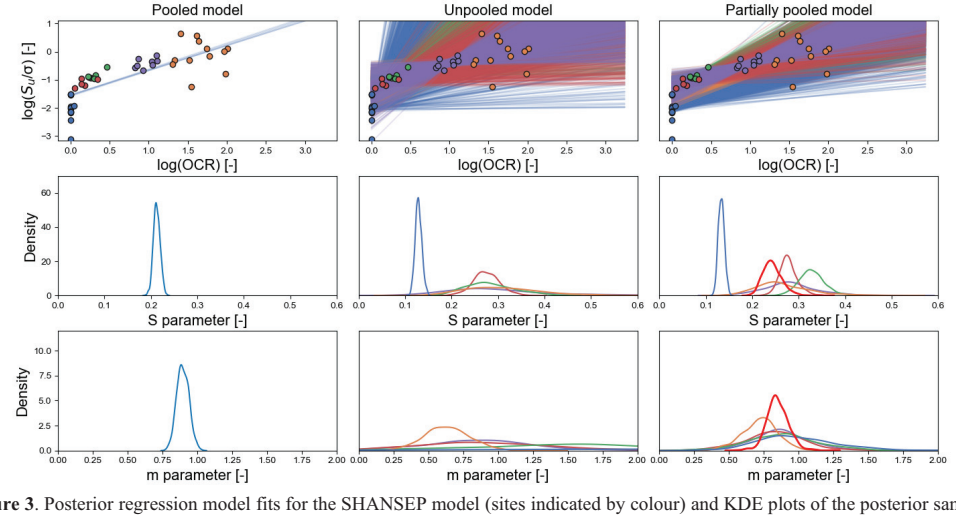

Figure 3. Posterior regression model fits for the SHANSEP model (sites indicated by colour) and KDE plots of the posterior samples for the pooled, unpooled and partially pooled models (only 5 sites displayed for clarity).

4. Impact of HBM in a geotechnical application

So far, the benefits of adopting HBM in deriving geotechnical parameters have been elaborated by analysing the structure of the statistical models and by comparing the resulting posteriors. This section aims to further investigate the advantages of employing HBM by assessing its impact on the failure probability of a geotechnical structure.

Specifically, the posteriors derived in the previous section (Fig. 2 and Fig. 3) are utilized in the failure probability assessment of the slope stability of an embankment. An artificial example is set up, representing a situation that is commonly met in practice: assessing the slope stability failure probability of an embankment. An existing sand embankment is assumed to be founded on a soft clay formation, the strength of which is characterized by a CPT and OCR profile. The embankment height is 4m and the slope has been built at an inclination of 1:2. On top of the embankment, a permanent load of 20kN/m^2 is placed. For simplicity, all model parameters, expect for the strength of the soft clay layer, are considered deterministic. The geotechnical model is built in D-Stability [9]. The failure probability of the slope is calculated using Monte Carlo Simulations (MCS).

The analysis aims to compare the failure probability results as estimated using the standard practice parameter derivation approaches and HBM. Moreover, it does so in the two different scenarios. In the first one, it is assumed that the embankment lies in one of the sites of the dataset. This scenario mimics the situation of having both in-situ and laboratory soil investigation data in the site of the embankment and Bayesian inference can lead to site-specific predictions. In this scenario, the standard approach predicts by performing OLS regression using only the data of the examined site. A site with strength values at the upper side of the strength range is selected. In the second scenario, the embankment lies at a new site where only in-situ soil investigation data is available. This scenario explores the ability of the HBM to make predictions using the population-level posteriors, whereas the standard approach predicts by performing OLS regression on the pooled dataset.

4.2 Probability of failure with the N_{kt} regression model inference

Since the N_{kt} regression model connects the CPT cone resistance directly to the undrained shear strength, soil is modelled with a failure criterion of undrained shear strength independent from the stress level. All analyses for the N_{kt} setting are performed with the same CPT profile, generated by Random Field Generation, as indicated in [10]. The mean cone resistance is 0.6 MPa and the Coefficient of Variation is 10%. The random field is characterized by a vertical scale of fluctuation of θ_v=1.0m. In combination with the calculation of the vertical effective stress, the corrected cone resistance (q_{net}) is estimated and used to estimate the undrained shear strength of the soil.

According to the standard approach, the S_u profile is generated by drawing samples from the relevant predictive distribution defined by the regression model. For the site-specific analysis, the regression is fit on the dataset of the site, while for the new site analysis the regression is solved for the entire database. For the site-specific analysis, the posterior predictive distribution of the site is used to draw S_u samples. At this point, it is supposed that the N_{kt} is horizontally homogenous in the proximity of the embankment. This assumption allows the sampling of each S_u profile from the posterior predictive distribution with a single random N_{kt} draw. When sampling for the new site analysis with the HBM, the sampling of the N_{kt} value is preceded by the sampling of the site mean and standard deviation from the population distributions.

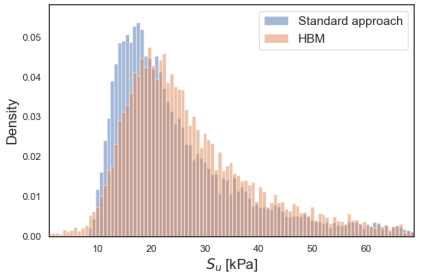

Figure 4. MCS samples for the N_{kt} setting reliability analysis, drawn according to the standard approach and the HBM.

The results are shown in Tab. 5. The standard method estimates a slope probability of failure (P_f) of 2% for the new site, while no failed realizations have been met in the site-specific analysis (likely implying a P_f < 1%). The HBM prediction leads to a P_f of 3% for the site-specific analysis, which is greater than that of the standard approach. A greater increase is found in the new site scenario, where the HBM calculates a P_f of 11%. In the N_{kt} setting, the regression coefficient posterior of the HBM in each scenario and the coefficient distribution derived by the OLS are centred in the same neighbourhood. However, in this scenario the HBM allows for greater uncertainty than the standard approach. Fig. 4 presents the histograms S_u samples drawn by the standard approach and the HBM. While the samples of the latter have a higher mean, they also exhibit greater variance. So, several HBM samples take values lower than the minimum of the standard approach samples. Since the latter shows no failed realizations, it is deduced that these samples generate the failures met in the MCS of the HBM. This greater variance of the HBM is attributed to its two-level structure of stochastic variables, as well as its probabilistic interpretation of the variables of the problem.

4.3 Probability of failure with the SHANSEP regression model inference

The second geotechnical setting employs the SHANSEP strength model and its HBM results in the slope failure probability analysis. The OCR profile is generated based on a uniform-over-depth POP value of 20 kPa. It should be noted that this geotechnical model is not equivalent to the previous one. No relationship has been accounted between the CPT cone resistance and the OCR value of the previous and current analysis respectively, so the soil material is not equally strong in the two settings.

As in the previous scenario, the undrained shear strength comes from the posterior predictive distribution of the HBM. Fig. 5 compares the mean regression model and the 90% credible intervals of the standard approach and the

HBM for the site-specific predictions. Due to the effect of the hyperparameters, which leads to well-defined posteriors of the regression coefficients, the credible interval of the HBM behaves better than that of the standard approach. Essentially, the credible interval of the latter implies that the model has predictive power only in the range of the dataset. In contrast, the opposite is true for the HBM, which has significant accuracy even far from the dataset. This behaviour was also verified by the performance of the HBM in the test metrics (section 3.4). Because the OCR profile of the geotechnical model spans a range of values greater than the one of the dataset, the HBM leads to less uncertain S_u samples that the standard approach.

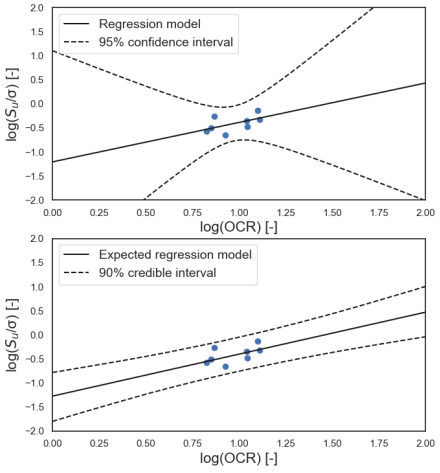

Figure 5. Mean regression model and 90% credible interval of the prediction accosting to the standard approach (top) and the HBM (bottom) for the site-specific prediction.

Table 5. Probability of failure per approach (column) and scenario (row) using the HBM posteriors in geotechnical settings N_{kt} and SHANSEP.

Approach / Scenario	Probability of failure Pf [%]			
	N_{kt} model		SHANSEP model	
	HBM	Standard	HBM	Standard
High strength site	3	< 1	2	46
New site	11	2	10	18

The probability of failure results are presented in Tab. 5. Similarly, to the behaviour exhibited by the unpooled model, the site-specific OLS regression for the standard approach leads to ill-defined distributions for the regression coefficients, which lack predictive power. So, the standard approach leads to high P_f (46%). On the other hand, the HBM can reach well-defined posteriors and attain considerable predictive prowess. The result of adopting HBM in the site-specific analysis in the reduction of P_f to 2%. For the new site analysis, the standard approach fits a regression model on the entire dataset. Much like the pooled model, the OLS approach leads to low coefficient uncertainty but high regression error. The strength prediction of this approach leads to a probability of failure of 18%. Contrarily, the HBM presents greater hyperparameter uncertainty but attains better flexibility, which in turn leads to lower regression error. Eventually, the HBM achieves a P_f of 10%. The new site prediction of the HBM has greater failure probability than the existing site because it includes greater uncertainty, but also due to the selected existing site showing strength values greater than the database average.

5. Conclusions

In the context of this paper, HBM has been combined with Robust Linear Regression to derive geotechnical parameters for two typical parameter estimation problems met in geotechnical practice.

Firstly, the HBM outcome posterior distributions, have been compared to the results of simpler Bayesian alternatives on a statistical level. Secondly, the impact of adopting HBM in parameter derivation is assessed on a geotechnical engineering level and compared with the standard practice approach. These comparisons are done for two geotechnical settings, the N_{kt} regression model and the SHANSEP model regression, linked to a slope stability analysis.

In the first geotechnical setting (N_{kt} model), and when comparing the HBM outcome, the HBM performs equally well to the unpooled model. However, the latter suffers in practice, as it lacks the derivation of population-level statistics. In the second geotechnical setting (SHANSEP model), the HBM outperforms both the pooled and unpooled models. Specifically, the power of the HBM in inferring the population distributions urges the site-specific posteriors to follow a global trend and guides them to explicit neighbourhoods, leading to well-behaved models. On the other hand, the unpooled model lacks this structure and exhibits ill behaviour, due to irregularities of the SHANSEP data. Notably, two regression coefficients are inferred for the SHANSEP model, while only one is inferred for the N_{kt} model. This hints that the HBM starts outperforming the unpooled model when the regression model becomes more complex. All in all, HBM is proven to be flexible enough to follow data patterns, is inherently prepared to avoid specific caveats and holds practical value, as it allows inference and prediction on site and population levels.

For the problem at hand, the HBM is structurally more uncertain than the unpooled and pooled models in an OLS regression scheme used in the standard approach. However, the HBM can lead to reduction of epistemic uncertainty and more accurate description of aleatory uncertainty, enough to compensate for its inherently greater uncertainty. This behaviour has been detected in the SHANSEP setting, which uses two regression coefficients and deals with more irregular data. In contrast, the HBM has led to greater uncertainty in the N_{kt} setting, which utilizes a simpler regression model.

To the end of comparing the impact of HBM estimations in a geotechnical application, the posteriors drawn are used in a slope stability failure probability analysis for an artificial example of an existing embankment. The analysis is performed for a site that already exist in the CLAY10 database, as well as for a new

site, whose samples are drawn from the population-level posterior.

In the N_{kt} setting, the HBM predictions have led to greater probabilities of failure than the standard approach. In this simple regression setting, the HBM does not outperform the OLS regression in flexibility and accuracy enough to compensate for its greater prediction uncertainty. On the other hand, adopting the HBM has been successful in drastically reducing the probability of failure for the SHANSEP setting, which entails a more complex regression model than the N_{kt} setting. In the Bayesian perspective, the HBM is still more credible than the OLS approach, as it provides a more thorough description of uncertainty, by modelling the parameters of the population level as stochastic variables.

In conclusion, HBM has exhibited considerable advantages for inference of geotechnical parameters in the two examined settings. Firstly, the hierarchical structure provides a twofold benefit in terms of statistical modelling; it allows for flexible models that achieve high prediction accuracy and drives models into deriving well-defined posteriors. Secondly, HBM aids on a practical level, due to its ability to infer population distributions and identify trends. Moreover, HBM can be easily adjusted to follow standard methods and produce results compatible with standard practice. Lastly, it has been demonstrated that adopting the HBM can lead to both lower and greater probabilities of failure than the ones estimated by the standard approach, depending on the complexity of the regression setting that is used to predict the soil strength for the geotechnical failure probability analysis.

References

[1] M. D. Hoffman and A. Gelman, "The No-U-Turn Sampler: Adaptively Setting Path Lengths," *Journal of Machine Learning Research ,* vol. 15, pp. 1351-1381, 2014.

[2] J. Salvatier, T. Wiecki and C. Fonnesbeck, *PyMC3,* 2016.

[3] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.

[4] M. de Koning, T. D. Y. F. Simanjuntak, D. G. Goeman, H. L. Bakker, J. K. Hasnoot and C. Bisschop, "Determination of SHANSEP parameters by laboratory tests," in *XVII European Conference on Soil Mechanics and Geotechnical Engineering*, Reykjavik, 2019.

[5] TC304, "TC304 Engineering Practice of Risk Assessment & Management," [Online]. Available: http://140.112.12.21/issmge/tc304.htm.

[6] M. Betancourt and M. Girolami, "Hamiltonian Monte Carlo for Hierarchical Models," in *Current Trends in Bayesian Methodology with Applications*, 2015.

[7] G. R. Terrell and D. W. Scott, "Variable kernel density estimation.," *The Annals of Statistics,* pp. 1236-1265, 1992.

[8] D. Lewandowski, D. Kurowicka and H. Joe, "Generating random correlation matrices based on vines and extended onion method," *Journal of Multivariate Analysis,* vol. 100, no. 9, pp. 1989-2001, 2009.

[9] Deltares, *D-Stability,* Delft: Deltares, 2019.

[10] A. P. van den Eijnden and M. Hicks, "Efficient subset simulation for evaluating the modes of improbable slope failure," *Computers and Geotechnics,* vol. 88, pp. 267-280, 2017.