

Research Article

Comparisons of Two Approaches for Geotechnical Model Calibration with Scarce Data

Yuanxin Lei , Huifen Liu , and Zhixiong Lu 

School of Transportation Civil Engineering and Architecture, Foshan University, Foshan 52800, Guangdong, China

Correspondence should be addressed to Huifen Liu; liuhuifen99@163.com

Received 8 November 2021; Revised 20 November 2021; Accepted 22 November 2021; Published 10 December 2021

Academic Editor: Yonghong Wang

Copyright © 2021 Yuanxin Lei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Geotechnical models are usually built upon assumptions and simplifications, inevitably resulting in discrepancies between model predictions and measurements. To enhance prediction accuracy, geotechnical models are typically calibrated against measurements by bringing in additional empirical or semiempirical correction terms. Different approaches have been used in the literature to determine the optimal values of empirical parameters in the correction terms. When measured data are abundant, calibration outcomes using different approaches can be expected to be practically the same. However, if measurements are scarce or limited, calibration outcomes could differ significantly, depending largely on the adopted calibration approach. In this study, we examine two most commonly used approaches for geotechnical model calibration in the literature, namely, (1) purely data-catering (PDC) approach, and (2) root mean squared error (RMSE) method. Here, the purely data-catering approach refers to selection of empirical parameter values that minimize coefficient of variation of model factor while maintains its mean value of one, based solely on measured data. A real case of calibrating the Federal Highway Administration (FHWA) simplified facing load model for design of soil nail walls is illustrated to thoroughly elaborate the differences in practical calibration and design outcomes using the two approaches under scarce data conditions.

1. Introduction

It has been well recognized that model uncertainty plays a key role in reliability-based design of geotechnical structures [1–5], as it is usually much larger than uncertainty associated with design parameters (e.g., soil cohesion, unit weight, and internal friction angle). Typically, geotechnical models need to be assessed and calibrated against measured or observed data before used for design. However, in many cases, measurements or observations that are available for model assessment and calibration are limited, mainly due to two reasons: first, obtaining in situ geotechnical data is costly and time consuming in general; and second, monitored data were always undervalued and thus not well collected and pooled, although in recent years, geotechnical engineers start to realize the value of data and make effort to make the best use of it [6–13].

Despite of the situation, assessment and calibration of geotechnical models using limited data are far better than doing nothing at all [3]. Usually, calibration of a geotechnical

model can be done by following two steps: (1) introduce an empirical or semiempirical correction term to a model and (2) determine the constants in the correction term according to certain criteria. Based on the calibration criteria, there are two methods that have been widely adopted for geotechnical model calibration in the literature. One is the purely data-catering (PDC) approach, and the other is the root mean squared error (RMSE) approach.

The PDC approach calibrates a model by adjusting the constants to satisfying two criteria: keeping the mean of bias equal to one while minimizing the coefficient of variation (COV) of bias. Here, bias is defined as the ratio of measured to the predicted value. Bias, also referred to as model bias or model factor elsewhere, is commonly treated as a random variable and used as an indicator for quantification of model accuracy. Previous geotechnical model calibration using the PDC approach can be seen in, e.g., Lin and Liu [14], Lin et al. [15], Lin et al. [16], Phoon and Kulhawy [17], Phoon and Tang [18], Tang and Phoon [19], Yuan et al. [20], and Yuan et al. [21].

On the other hand, the RMSE approach determines the constants as the set that minimizes the root mean squared error between measurements and predictions. Common examples are geotechnical models developed using the response surface method and machine learning methods, e.g., Bathurst and Yu [22], Lin et al. [16], Liu et al. [23], Yu and Bathurst [24], Zhang et al. [25], Zhang et al. [26], Zhang et al. [27], and Zhang et al. [28]. Obviously, these two calibration approaches are not equivalent, especially when the data for calibration are scarce. This will result in different calibrated values for the constants, which in turn lead to different geotechnical design outcomes. Discussion on the difference of these two approaches will be provided later in this study.

While being extensively used, influences of adoption of the two approaches on geotechnical calibration outcomes and the consequences are not yet thoroughly examined. To fill the gap, this study is focused on investigating the differences in both calibration and practical design outcomes that are resulted by using the above two calibration approaches. To allow comparing model competence from a third angle, the Bayesian information criterion (BIC) is employed [29]. A case study of calibrating the default Federal Highway Administration (FHWA) facing load model for facing design of soil nail walls is shown to elaborate the influences.

2. Approaches for Model Calibration and Ranking

This section introduces in detail the PDC and RMSE approaches for calibration of geotechnical models against observed data. The commonalities and differences of the two approaches are discussed. A likelihood-based model ranking method called Bayesian information criterion (BIC) is also introduced, which will be used later to quantify the competences of the calibrated models.

2.1. Purely Data-Catering Approach. Suppose $y = f(x)$ be the geotechnical model to be calibrated. Here, y is the model output which is taken to be a scalar for simplicity; and x is the model input parameter vector, $x = (x_1, x_2, \dots, x_n)$. Note that the input parameter vector x is the same across different design scenarios, while the values of its elements could vary. For convenient, we denote y_k and x_k for design scenario k .

Let $D = (d_1, d_2, \dots, d_m)$ be m observed or measured values for y from m real cases, and λ be the model factor (which is a random variable) for $y = f(x)$. Then, based on the method of moment, the sample mean and sample standard deviation of λ , denoted as $\bar{\lambda}$ and s_λ , respectively, can be computed as

$$\begin{aligned} \bar{\lambda} &= \frac{1}{m} \sum_{k=1}^{k=m} \frac{d_k}{y_k}, \\ s_\lambda &= \sqrt{\frac{1}{m-1} \sum_{k=1}^{k=m} \left(\frac{d_k}{y_k} - \bar{\lambda} \right)^2}. \end{aligned} \quad (1)$$

The purely data-catering (PDC) approach assumes that a model can be adequately (indeed, compromisingly) calibrated in terms of λ and s_λ that are obtained by the method of moment. The PDC approach introduces an empirical correction term, M , to the original model for calibration purposes, where M is a function of x and C , with C being a vector of empirical constants to be determined using D . In general, the correction term can be written as $M = f_M(x; C)$. As such, the calibrated model can be expressed as

$$y' = M \times f(x) = f_M(x; C) \times f(x), \quad (2)$$

where y' is the model output after calibration. The sample mean and sample standard deviation of the model factor for the calibrated model, denoted as λ' , can then be calculated as

$$\bar{\lambda}' = \frac{1}{m} \sum_{k=1}^{k=m} \frac{d_k}{y'_k}, \quad (3)$$

$$s_{\lambda'} = \sqrt{\frac{1}{m-1} \sum_{k=1}^{k=m} \left(\frac{d_k}{y'_k} - \bar{\lambda}' \right)^2}. \quad (4)$$

The calibration principles of the PDC approach are to adjust the values of the empirical constants in C until simultaneously satisfying two criteria: (1) $\bar{\lambda}'$ is equal to 1, and (2) $s_{\lambda'}$ is minimized. Evidently, when the data available for calibration are limited, i.e., m in equations (3) and (4) is not sufficiently large, both $\bar{\lambda}'$ and $s_{\lambda'}$ could be significantly influenced by m and D , resulting in unstable calibration outcomes.

2.2. Root Mean Squared Error (RMSE) Approach. This method assesses the accuracy of a model by computing its root mean squared error (RMSE) between model predictions (i.e., y_k or y'_k) and measured (true) values (d_k). The RMSE can be computed as

$$r = \sqrt{\frac{1}{m} \sum_{k=1}^{k=m} (y_k - d_k)^2}, \quad (5)$$

for the original model, i.e., $y = f(x)$, and

$$r' = \sqrt{\frac{1}{m} \sum_{k=1}^{k=m} (y'_k - d_k)^2}, \quad (6)$$

for the calibrated (corrected) model, i.e., $y' = f_M(x; C) \times f(x)$.

The calibration principle of the RMSE method is to select C that minimizes r' . Note that this method does not necessarily result in $\bar{\lambda}' = 1$ with minimal $\text{COV}'_{\lambda'}$, where $\text{COV}'_{\lambda'} = s_{\lambda'}/\bar{\lambda}'$ is the coefficient of variation of λ' . Although the RMSE method has been widely used for geotechnical model development and calibrations, it does not provide an intuitive impression of model accuracy.

2.3. Model Competence Ranking. The Bayesian information criterion (BIC) is a relative measure of goodness-of-fit among models given observed data and has been widely used in ranking model competence [29, 30]. The BIC is computed as

$$BIC = -2 \ln(L_{\max}) + \kappa \ln(n), \quad (7)$$

where $\ln(L_{\max})$ is the log of the maximum likelihood, κ is the number of model parameters, and n is the number of data points.

Typically, the maximum value of the likelihood function (L) and the corresponding maximum likelihood estimators can be found using numerical methods. Technical descriptions of the maximum likelihood estimation method, e.g., the construction of a likelihood function, are not provided here for brevity. Interested readers are directed to, e.g., Juang et al. [31] and Lin and Liu [14] for more details.

The criterion simply states that the smaller the BIC value of a fitting model, the better the model captures the observations. It should be emphasized that the absolute value of the BIC itself is meaningless in terms of model competence; only the difference between BICs helps ranking the models.

2.4. Discussion. Although the above two approaches have been widely adopted for geotechnical model calibrations based on observed data, there are some fundamental differences in calibration outcomes and interpretations, as given in Table 1.

The PDC approach uses two indicators to jointly describe the model accuracy, i.e., $\bar{\lambda}'$ and s'_λ ; whereas, the RMSE approach uses only one indicator, r' . Typically, $\bar{\lambda}'$ is interpreted as an indicator for on-average accuracy, while s'_λ is taken as an indicator for dispersive accuracy. The advantage of such an accuracy assessment scheme is that it provides an immediate and general idea of the performance of a model. For example, $\bar{\lambda}' = 0.90$ and $\text{COV}'_\lambda = s'_\lambda/\bar{\lambda}' = 0.30$ suggest that overall model predictions are 10% larger than the corresponding observations, while the dispersion in prediction accuracy is 30%. Clearly, if a model is perfect, then it would have $\bar{\lambda}' = 1$ and $\text{COV}'_\lambda = 0$. The disadvantage is probably the lack of ability to compare accuracies among models. This can be easily seen for two models, for example, A and B, where A has $\bar{\lambda}' = 0.90$ and $\text{COV}'_\lambda = 0.30$ and B has $\bar{\lambda}' = 0.80$ and $\text{COV}'_\lambda = 0.20$. In such a case, model A has a better on-average accuracy, but its prediction is more dispersive; whereas, model B has a less on-average accuracy, but the prediction spreads less. Hence, it is difficult to directly determine which model is more accurate, if without any further analyses.

For the RMSE approach, the smaller the r' , the better the model accuracy. For a perfect model, $r' = 0$. As this method uses a single index for model assessment, accuracy comparison among models is straightforward. However, this approach does not provide an intuitive sense of accuracy of the model itself.

Another difference is the criteria set for calibration. The PDC approach minimizes s'_λ or COV'_λ conditioned on $\bar{\lambda}' = 1$. A mean of λ' of one represents that the calibrated model is

TABLE 1: Comparisons of the two model calibration approaches.

Approach	Accuracy indicator	Calibration criterion
PDC	$\bar{\lambda}'$ and s'_λ	Maintain $\bar{\lambda}' = 1$ and minimize s'_λ
RMSE	r'	Minimize r'

unbiased (on average, within the context of observations D). The objective of the RMSE approach is to minimize r' to obtain the best accuracy. As has been pointed out earlier, usually this is not necessarily equivalent to the criteria for the PDC approach.

Generally, both PDC and RMSE approaches can only handle uncensored data. If observation data are censored, then more robust approaches such as the maximum likelihood method or Bayesian inference technique can be employed. However, to do so, the maximum likelihood method and Bayesian approach would have to first assume the probability distribution of λ' so as to construct the likelihood function L . Therefore, the estimators by these approaches are conditioned on the distribution of λ' . For PDC and RMSE approaches, $\bar{\lambda}'$ and s'_λ and r' are computed without any assumptions of distribution of λ' .

Last, it is pointed out that while each measurement or observation is equally weighted in the PDC calibration approach, it is not the case in the RMSE calibration approach. For the RMSE approach, measurements with large values weigh much more than those with small values in the calibration process.

3. Case Study

This section presents a case study to elaborate the difference in model calibration outcomes using the two approaches. The example is to calibrate the default Federal Highway Administration (FHWA) simplified model for computation of facing loads of soil nail walls using a total of 23 measured data. These data were collected by Liu et al. [32] from the literature. They corresponded to facing loads monitored during or at completion of wall constructions. Hence, they should be interpreted as “short-term” facing loads.

In the following, the measured facing load database established by Liu et al. [32] is first introduced, followed by a brief review of the default FHWA simplified facing load model. Section 3.3 presents the calibration results along with comparisons and discussion. Note that calibration of the FHWA facing load model has been done by Liu et al. [32] using the PDC approach. Section 3.4 shows how would the selection of calibration approaches affect the practical designs of facing of soil nail walls.

3.1. Database of Short-Term Facing Loads. Figure 1 shows the side and front views of the facing of a typical soil nail wall. Nails are structurally connected to the facing at their heads. As the wall deforms, lateral active earth pressures act onto the facing, which are then transferred to nails due to the nail-facing connections. In equilibrium state, a nail is responsible for the lateral earth pressure within a tributary area where the nail head centers. The product of the earth pressure and

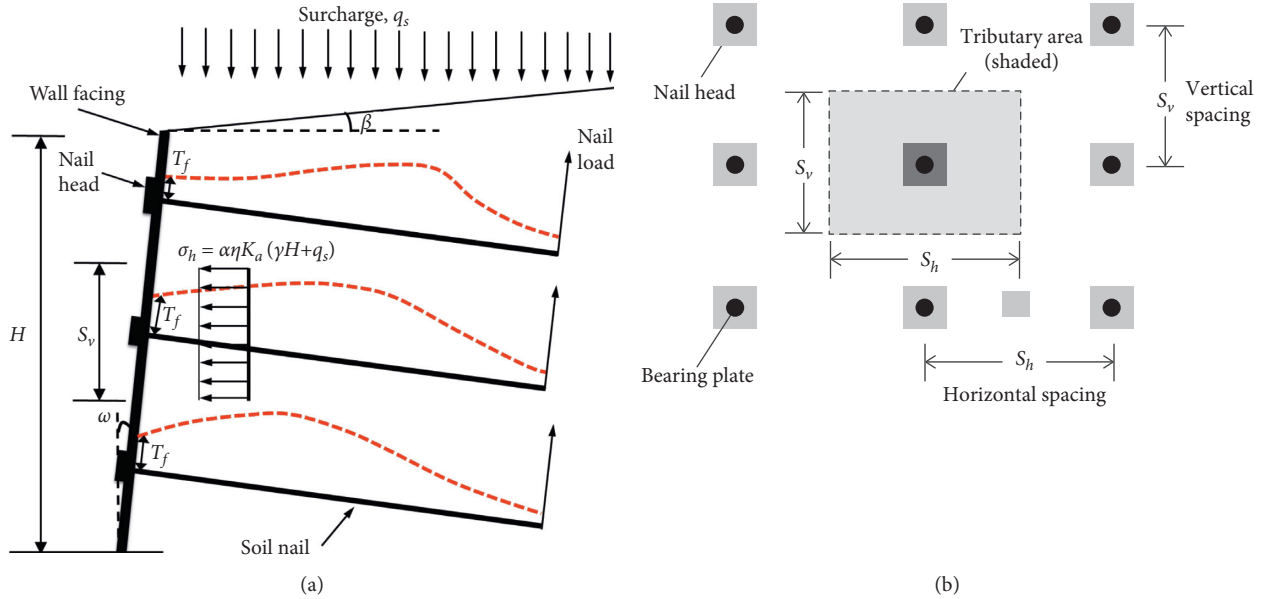


FIGURE 1: Facing of a typical soil nail wall: (a) side view; (b) front view.

the tributary area is referred to as the nail head tensile load or facing load (Figure 1(a)) in this study. Here, the tributary area is the product of horizontal and vertical nail spacing (Figure 1(b)).

Liu et al. [32] developed a database containing measured long-term and short-term facing loads. The short-term load data are extracted and briefly reviewed here for the reader's convenience to follow. They collected in total 31 short-term load data; however, 8 were identified as questionable data and thus filtered from analyses. Table 2 provides the wall geometry, soil properties, facing type, and nail spacing for soil nail walls where the remaining 23 data were from.

The data were collected from five wall sections, ranging from 4 to 12 m high. All walls had vertical or steep facings and horizontal back slopes. Four walls were in cohesionless soils, while one was in cohesive soil. Two walls were subjected to surcharge, in addition to soil self-weights. The facings were constructed with shotcrete or concrete panels. Nail spacings were set between 1 and 2 m, which are typical. Readers are directed to Liu et al. [32] for detailed description of the collected facing load data.

3.2. FHWA Facing Load Models. The facing of a soil nail wall, as shown in Figure 1, can be simplified as a continuous two-way slab. According to the FHWA soil nail wall design manuals [33], under working conditions, the maximum facing load due to the lateral active earth pressure, T_f , can be calculated as

$$T_f = \alpha \eta K_a (\gamma H + q_s) S_h S_v, \quad (8)$$

where α is the empirical spacing factor expressed as $\alpha = 0.6 + 0.2(S_{\max} - 1)$; S_{\max} (unit: m) is the larger of horizontal and vertical nail spacing S_h and S_v , respectively; η is the empirical depth factor expressed as $\eta = 1.25 h/H + 0.5$ for $0 < h/H \leq 0.2$, $\eta = 0.75$ for $0.2 < h/H \leq 0.7$, and

$\eta = 2.03 - 1.83 h/H$ for $0.7 < h/H \leq 1$, where h and H are the depth and wall height, respectively; K_a , γ , and q_s are the Coulomb active earth pressure coefficient, soil unit weight, and surcharge, respectively.

The measured facing loads (T_m) are plotted against the corresponding predicted T_f using equation (8), as shown in Figure 2(a). The data points in the figure appear to be two clusters: one around the 1:1 correspondence line and the other below the line. This suggests that the current default FHWA facing load model is conservative as generally it would overestimate the maximum facing loads. This observation is confirmed by computing the sample mean ($\bar{\lambda}$) and sample COV (COV_λ) of λ of equation (8), which are $\bar{\lambda} = 0.77$ and $COV_\lambda = 0.669$. Hence, on average, equation (8) overpredicts the maximum facing loads by 23%, and the predictions are highly dispersive, according to the ranking scheme proposed by Phoon and Tang [18]. Furthermore, Figure 2(b) shows that λ tends to decrease as T_f increases. The dependency is quantitatively confirmed at a level of significance of 0.05 by Spearman's rank correlation test results that are also given in the figure. Such a dependency is undesirable, and its effects on reliability-based geotechnical design have been investigated by Lin and Bathurst [34]. As a result, it is necessary to perform model calibration for equation (8) for accuracy improvements.

To identify the sources of model errors and the abovementioned dependency, λ are plotted against each input parameter of equation (8), as shown in Figure 3. Spearman's rank correlation test results show that λ are statistically correlated to α , K_a , and $S_h S_v$ at a significance level of 0.05. Therefore, a correction term which is a function of these three parameters can be introduced to the equation for calibration, i.e., $M = f(\alpha, K_a, S_h S_v)$. However, since α and $S_h S_v$ are highly correlated, only $S_h S_v$ is kept, while α is removed from M formulation for simplification. Moreover, K_a is removed from M to further simplify the calibration as

TABLE 2: Summary of wall geometry, soil properties, facing type, and nail spacing for soil nail walls.

Wall	Soil type	Wall geometry			Soil strength properties			q_s (kPa)	Facing type	Nail spacing	
		H (m)	ω ($^\circ$)	β ($^\circ$)	ϕ ($^\circ$)	c (kPa)	γ (kN/m 3)			S_h	S_v
W1	Heterogenous SM	9.2	0	0	36.5	18.5	16.3	0	Shotcrete	1.85	1.85
W2	Glacial tills	12	10	0	32	0	20	22.5	Shotcrete	1.5–1.9	0.9–1.55
W3	Fill, silt, sand	7.6	0	0	38	0	19.6	0	N/A	1.83	1.22–1.83
W4	Gravelly silty sand	4	0	0	35	0	19.6	0	Concrete panel	1.0	1.05
W5	Gravelly silty sand	5	0	0	35	0	19.6	127	Concrete panel	1.0	1.05

H , wall height; ω , face batter angle; β , back slope angle; ϕ , soil friction angle; c , soil cohesion; γ , soil unit weight; q_s , surcharge; S_h , horizontal nail spacing; S_v , vertical nail spacing.

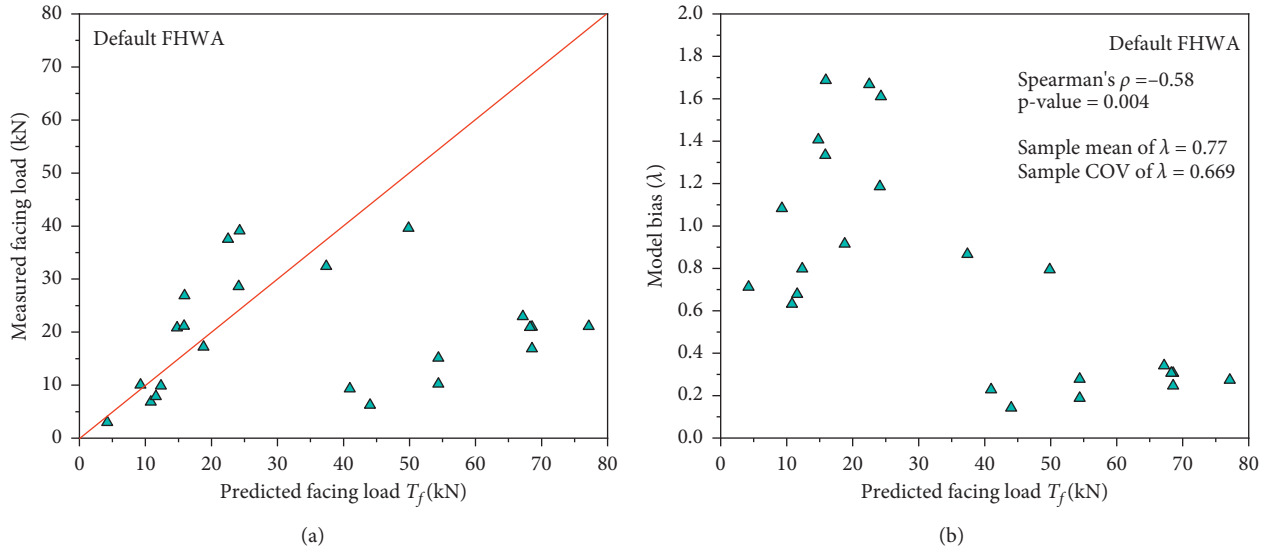


FIGURE 2: Analyses of accuracy of the default FHWA simplified facing load model: (a) measured versus predicted facing load; (b) model bias versus predicted facing load.

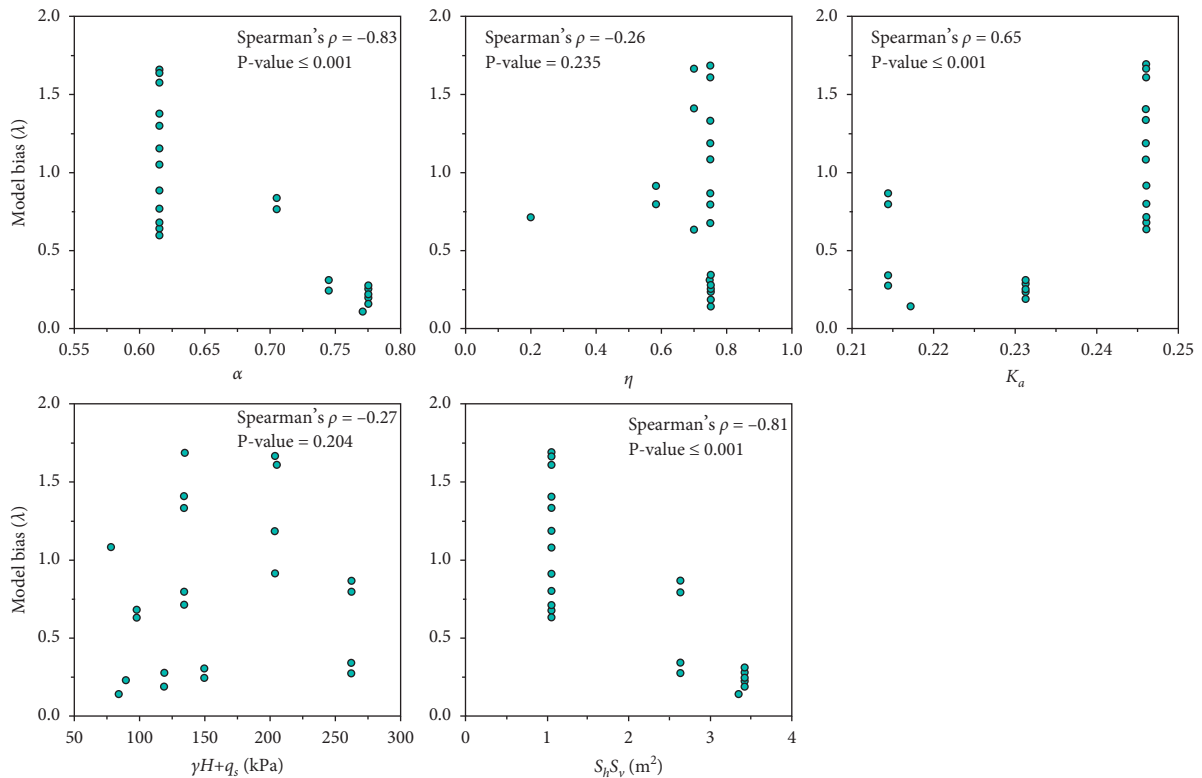


FIGURE 3: Plots of model bias λ for the default FHWA facing load model against various input parameters.

the focus of this study is on comparison of calibration approaches rather than model development. Last, a power form expression is assumed for M , and thus, equation (8) becomes

$$T_f = M\alpha\eta K_a (\gamma H + q_s) S_h S_v = a \left(\frac{S_h S_v}{A_t} \right)^b \alpha\eta K_a (\gamma H + q_s) S_h S_v, \quad (9)$$

where a and b are the empirical constants to be determined using the 23 measured facing load data; and $A_t = 2.25 \text{ m}^2$ is the typical tributary area used to make M dimensionless. The PDC and RMSE approaches discussed earlier in this study are then used to determine the values of a and b . The results are shown and discussed in the next section.

3.3. Analysis Results. Calibration of equation (9) is carried out in this section. The accuracy of the calibrated model is then compared based on sample mean and sample COV and also root mean squared errors (RMSEs). After that, the accuracy is rescrutinized from a maximum likelihood perspective, which helps understanding the calibration outcomes.

3.3.1. Comparisons between PDC and RMSE Approaches. Table 3 provides the calibration outcomes for equation (9) using the two approaches. By using the PDC approach, the constants in equation (9) are determined as $a = 0.4964$ and $b = -1.0677$, which correspond to $\bar{\lambda}' = 1.00$ and a minimal $\text{COV}'_{\lambda} = 0.432$. The RMSE for this case is computed to be 38.3618. On the other hand, by the RMSE approach, the optimal a and b values are found to be 0.4866 and -1.3118 , respectively. The minimal RMSE is 35.9635. The sample mean and sample COV for this case are $\bar{\lambda}' = 0.96$ and $\text{COV}'_{\lambda} = 0.452$. Several interesting observations can be made from these results.

First, these two approaches are not practically equivalent from the perspective of model calibration when the data volume available for the calibration is small. The resulted a values are close, i.e., 0.4964 versus 0.4866; whereas, the b values are significantly different, i.e., -1.0677 versus -1.3118 .

Second, based on bias statistics of equation (9), the PDC approach appears to be superior as its results are unbiased on average and less dispersive compared to those by the RMSE approach (i.e., 4% overestimation and 2% more dispersive). Interestingly, if comparing the RMSEs, one would easily reach a reverse conclusion that the RMSE approach is better than the PDC approach as it gives less RMSE, 35.9635 against 38.3618. Therefore, it is difficult to judge which one is better if based merely on the results given in Table 3. The comparison should be made from a third angle, which will be discussed in the next subsection.

Last, both approaches are effective and efficient in model calibration for accuracy enhancement. The essence of calibration is to move the data points towards the 1:1 correspondence line in general, as shown in Figure 4. The difference is that λ' by the PDC approach seems to be more skewed, while those by the RMSE approach appear to be

more uniform. Despite of this, λ' by the two approaches follow closely along the 1:1 correspondence line, and thus, the two methods in general do not lead to fatally different outcomes.

3.3.2. Rescrutinization from a Maximum Likelihood Perspective. The accuracies of the three facing load models, i.e., default FHWA, calibrated FHWA by PDC, and calibrated FHWA by RMSE, are reassessed in this section using the maximum likelihood method. The cumulative distributions of λ and λ' are shown in Figure 5. Kolmogorov-Smirnov tests are applied to the three bias datasets, and the results show that all the datasets can be considered as both normally and log-normally distributed at a significance level of 0.05. Hence, the maximum likelihood estimation (MLE) is carried out assuming normal and log-normal λ and λ' . The estimation outcomes are given in Table 4.

For both cases, the estimated means by MLE are practically the same as the sample means; while, the estimated COVs by MLE are slightly less than the sample COVs, i.e., about 2–4% which are practically neglectable. However, for the default model case (i.e., equation (9)), the estimated bias COV by MLE assuming log-normal λ is much larger than the sample COV or that assuming normal λ , i.e., about 20% higher.

For the normal case, the computed maximum loglikelihood values $\ln(L_{\max})$ are -12.8317 and -12.9868 for the PDC and RMSE approaches, respectively. As for both models, the number of parameters is $\kappa = 2$, and the number of data points is $n = 23$; by using equation (7), the BIC values are correspondingly calculated to be 31.93 and 32.24. This means that, if λ' is a normal random variable, then in this case, the PDC calibration approach is better than the RMSE approach as the BIC value for the former is smaller. On the other hand, $\ln(L_{\max})$ are -9.5680 if using the PDC approach and -8.6353 if using the RMSE approach. The corresponding BIC values are 25.41 and 23.54. This suggests that if λ' is a log-normal random variable, then model calibration using the RMSE approach is preferable. For both approaches, the $\ln(L_{\max})$ values for the log-normal case are always less than those for the normal case; hence, it can be said that λ' is more likely a log-normal random variable. If comparing the four $\ln(L_{\max})$ values, one would conclude that the RMSE-log-normal scenario is the best one as its $\ln(L_{\max})$ is the smallest one.

3.4. Practical Influences. Analyses presented earlier show how the selection of a calibration approach would affect the calibration outcomes, i.e., $C = (a, b)$. In this part, we investigate the influences of the calibration outcomes on practical designs of facing of soil nail walls. The facing design must ensure adequate margins of safety against various limit states, including facing flexural, punching shear, and headed-studs tensile failures. For illustration purposes, here we only consider the facing flexural limit state, which requires estimation of the maximum facing load and the ultimate facing flexure capacity. The primary design parameter for this limit state is the reinforcement ratio cross-sectional area per unit width at the nail head

TABLE 3: Summary of calibration outcomes of equation (10) using the PDC and RMSE approaches.

Model	Approach	Constant in C		Model bias, λ or λ'		RMSE
		a	b	Sample mean	Sample COV	
Default FHWA	N/A	1	0	0.77	0.669	140.7869
Calibrated FHWA	PDC	0.4964	-1.0677	1.00	0.432	38.3618
	RMSE	0.4866	-1.3118	0.96	0.452	35.9635

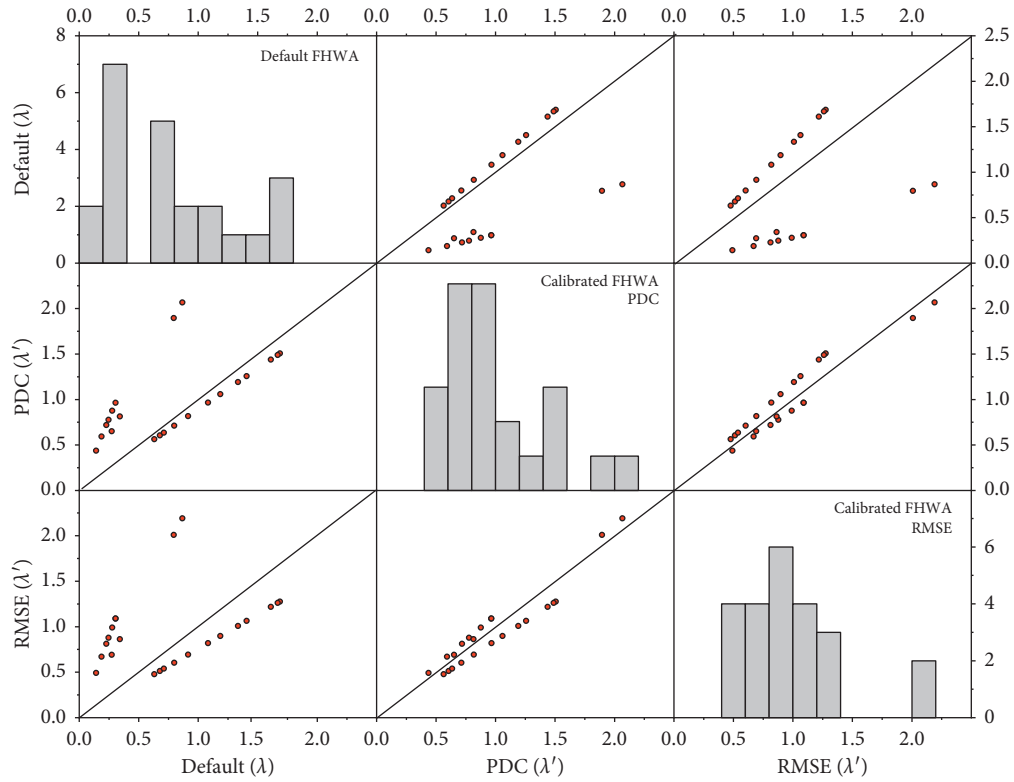


FIGURE 4: Histograms and comparisons of model biases λ and λ' .

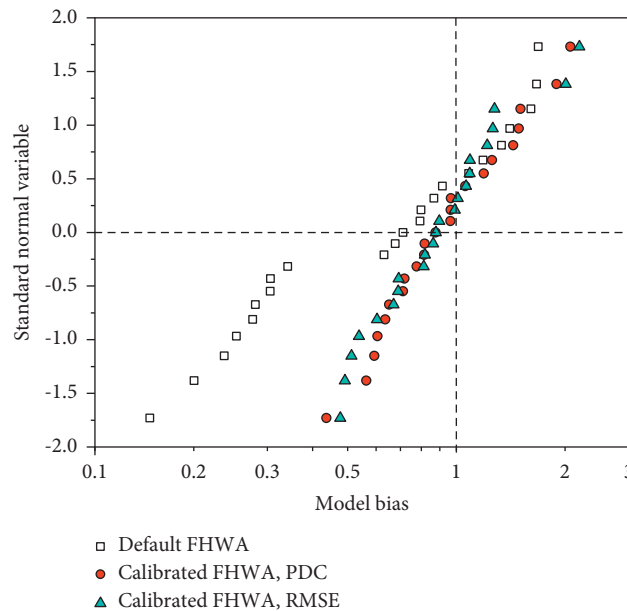


FIGURE 5: Cumulative distributions of model biases of the default and calibrated FHWA facing load models.

TABLE 4: Estimated statistics of model bias, λ , or λ' using the maximum likelihood method.

Model	Approach	Normal			Log-normal		
		Mean	COV	$\ln(L_{\max})$	Mean	COV	$\ln(L_{\max})$
Default FHWA	N/A	0.77	0.654	-16.8202	0.79	0.888	-14.3902
Calibrated FHWA	PDC	1.00	0.423	-12.8317	1.00	0.414	-9.5680
	RMSE	0.96	0.442	-12.9868	0.96	0.414	-8.6353

and at midspan. The influences of the two calibration approaches are assessed by using C to compute the maximum facing loads and to determine the reinforcement ratio.

The example wall used for analysis is 10 m in height with a horizontal back slope ($\beta = 0^\circ$) and a vertical facing ($\omega = 0^\circ$). The means of the soil strength parameters are assumed to be $c = 0$ kPa, $\phi = 30^\circ$, and $\gamma = 18$ kN/m³. The COVs are taken as 0.15 for ϕ and 0.075 for γ . Nails are spaced at 1.2 m horizontally and vertically.

3.4.1. On Estimation of Facing Loads. Both equations (9) and (10) are used to compute the nominal facing loads (T_f) along depth (h/H). The results are shown in Figure 6. Visually, T_f along h/H is a trapezoidal shape with larger values in the middle and smaller values at the top and bottom of the wall. The default FHWA model gives the highest facing loads, for example, at $h/H = 0.5$, the nominal T_f is about 42 kN; while, the calibrated FHWA model by the PDC approach gives the smallest T_f , which is about 33 kN. The difference is over 20%. The T_f by the calibrated FHWA model by the RMSE approach is about 36 kN. The difference in T_f using the two approaches is about 8%.

Figure 7 shows the distributions of facing loads corrected by biases, λT_f or $\lambda' T_f$, where the statistics of λ and λ' are estimated by the method of moment (referred to as sample bias) and maximum likelihood method (referred to as MLE bias). Visually, the distributions of $\lambda' T_f$ by PDC and RMSE approaches are highly similar, regardless of the methods used to compute bias statistics. Compared to the calibrated FHWA cases, the distributions of λT_f for the default FHWA case move leftward, meaning that λT_f overall is smaller than $\lambda' T_f$. There is a noticeable difference in λT_f distributions by sample bias and MLE bias, mainly due to the large difference in COV of λT_f , as given in Table 5, which summarizes computed means and COVs of λT_f and $\lambda' T_f$ at $h/H = 0.5$ using different calibration approaches and bias estimation approaches. On the other hand, λT_f for the default FHWA case has longer tails than those for the calibrated FHWA cases; obviously, this is due to the much higher COVs of λT_f , compared to those for $\lambda' T_f$. Last, it is observed that, in general, the differences in the means and COVs of $\lambda' T_f$ based on RMSE and PDC approaches are practically insignificant, as given in Table 5.

3.4.2. On Facing Design. Consider identical reinforcement cross-sectional area per unit width at the nail head and at midspan ($a_n + a_m$) horizontally and vertically, with $S_h = S_v$, the facing flexural capacity, R_F , calculated as [33]

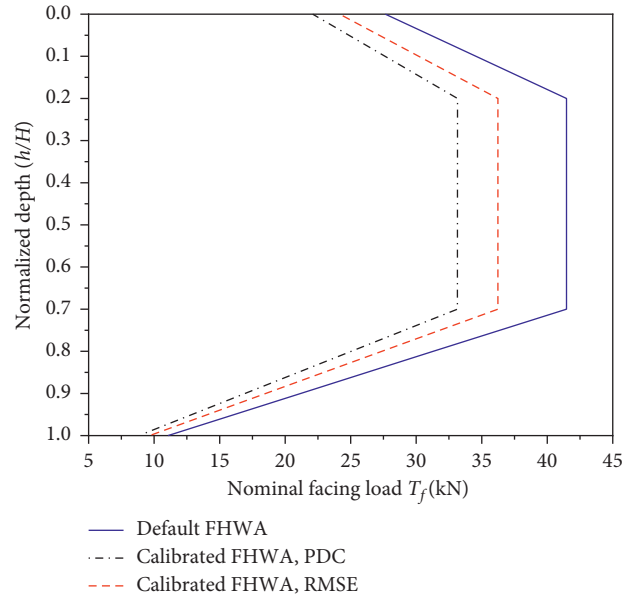


FIGURE 6: Predicted (nominal) maximum facing loads along depth using the three FHWA models.

$$R_F = \frac{C_F}{265} (a_n + a_m) t f_y. \quad (10)$$

where C_F is the empirical factor accounting for the non-uniformity of soil pressures behind the facing and is equal to 2.0, 1.5, and 1.0 for temporary walls when the facing thickness is $t = 100, 150,$ and 200 mm [35], and f_y is the reinforcement tensile yield strength. In this example, the facing thickness is taken as $t = 100$ mm, and thus, $C_F = 2.0$. The nominal value of f_y is taken as 414 MPa. The reinforcement area ($a_n + a_m$) is the main design parameter to be determined given a target margin of safety (e.g., factor of safety or reliability index). The design factor of safety for this limit state can be calculated as

$$F_S = \frac{R_F}{T_f}. \quad (11)$$

The performance function, g_F , can be written as

$$g_F = \lambda_R R_F - \lambda_L T_f. \quad (12)$$

where λ_R and λ_L are the model factors (model factors (biases) for R_F and T_f , respectively. In this study, λ_L is λ if using the default FHWA model and λ' if using the calibrated FHWA models. λ_R is taken as a log-normal random variable with mean of 1.1 and COV of 0.1 [36, 37].

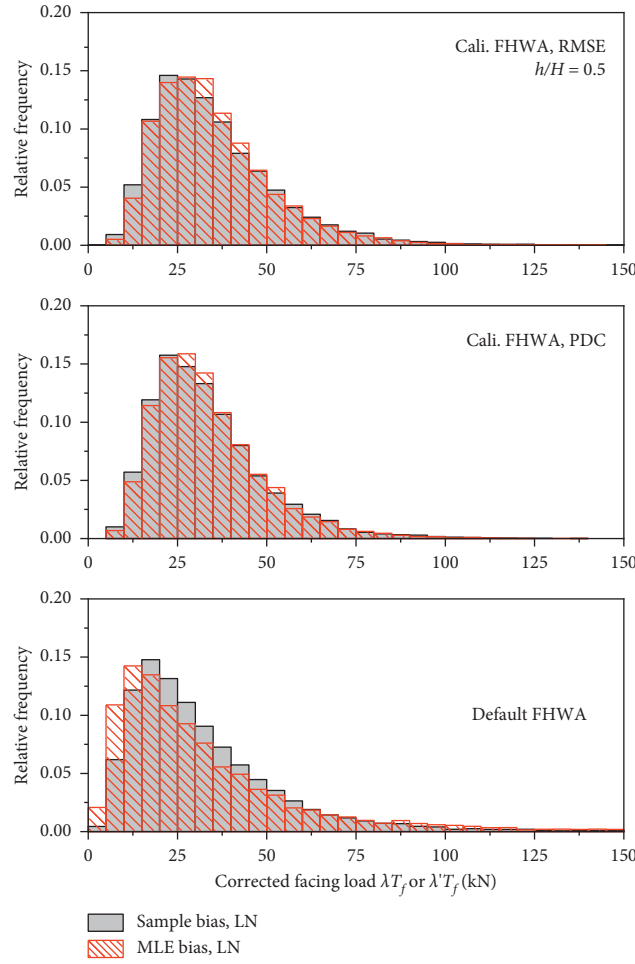


FIGURE 7: Histograms of facing load at $h/H = 0.5$ corrected by different biases, λT_f or $\lambda' T_f$.

TABLE 5: Summary of computed means and COVs of λT_f and $\lambda' T_f$ at $h/H = 0.5$ using different calibration approaches and bias estimation approaches.

Bias statistics	λT_f by default FHWA		$\lambda' T_f$ by calibrated FHWA, PDC		$\lambda' T_f$ by calibrated FHWA, RMSE	
	Mean (kN)	COV	Mean (kN)	COV	Mean (kN)	COV
By sample	32.21	0.700	33.39	0.478	35.21	0.500
By MLE	33.11	0.945	33.45	0.457	35.19	0.464

Table 6 provides the design outcomes of $(a_n + a_m)$ using the deterministic approach where the margin of safety is taken as $FS = 2.0$ and using the reliability approach where the margin of safety is taken as $\beta_T = 3.50$. Here, β_T is the target reliability index, and $\beta_T = 3.50$ roughly corresponds to a probability of failure of $1/5000$. Note that these FS and β_T values are consistent with those recommended in the FHWA soil nail wall design manual by Lazarte et al. [33] and Aashto [38]. Based on the deterministic approach (equation (11)), using the FHWA facing load model calibrated by the PDC approach gives the least $((a_n + a_m))$, which is $212 \text{ mm}^2/\text{m}$; whereas, the default FHWA model gives the highest value, i.e., $266 \text{ mm}^2/\text{m}$. The difference is about 20%.

However, based on the reliability approach, the computed $(a_n + a_m)$ values using the default FHWA model are

much larger than those using the calibrated FHWA models, i.e., 730 versus 446 and 492 and 1135 versus 430 and 447. The difference is about 40–60%. The design outcomes by the calibrated FHWA models by PDC and RMSE approaches are more or less similar, albeit those by the RMSE approach are slightly higher.

Figure 8 shows the $(a_n + a_m)$ values with respect to FS ranging from 1 to 5 and β_T ranging from 2 to 4 using the default and calibrated FHWA facing load models. It confirms that the difference in $(a_n + a_m)$ is insignificant between the two calibrated FHWA models; both are much less than those obtained by the default model. This highlights the importance of performing geotechnical model calibration while the influence of selection of the model calibration approach is secondary. Tables 5 and 6 and Figure 8 together

TABLE 6: Comparisons of reinforcement area per unit width for facing design against flexural failures using default and calibrated FHWA simplified facing load models.

Approach	Reinforcement area per unit width, $(a_n + a_m)$ (mm^2/m)		
	$FS = 2.0$	$\beta_T = 3.50$, sample bias	$\beta_T = 3.50$, MLE bias*
Default FHWA	266	730	1135
Cali. FHWA, PDC	212	446	430
Cali. FHWA, RMSE	232	492	447

* Assuming log-normal distribution for bias.

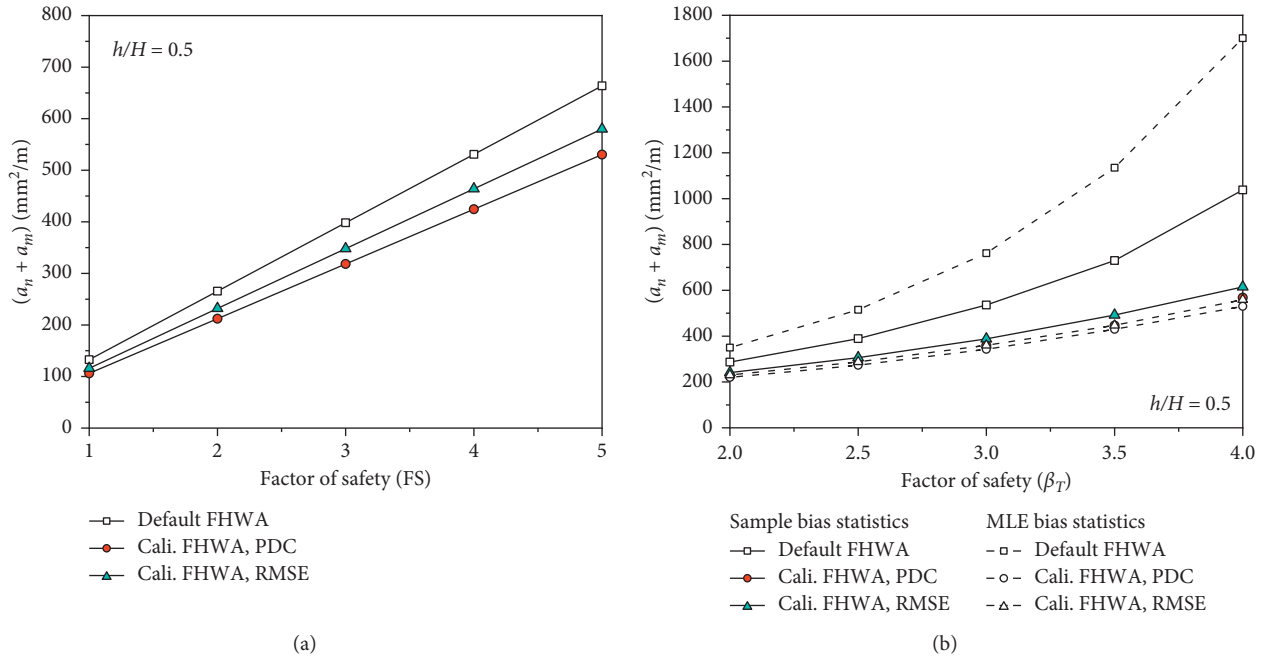


FIGURE 8: Plots of design outcomes of $(a_n + a_m)$ against FS ranging from 1 to 5 and β_T ranging from 2 to 4.

suggest that adopting the PDC and RMSE approaches for geotechnical model calibration does not result in a fatal difference on practical design outcomes, even under scarce data conditions.

4. Conclusions

Calibrations of geotechnical models in many cases have to be carried out with scarce data. This study examines two approaches that have been widely adopted for geotechnical model calibration in the literature, namely, pure data-catering (PDC) approach and root mean squared error (RMSE) approach. The PDC approach calibrates a model by adhering to two criteria: maintaining mean of model bias of one while minimizing COV of model bias, where model bias is defined as the ratio measured to the predicted value. The RMSE approach calibrates a model by minimizing the root mean squared error between measured and predicted values.

A case study is presented to elaborate the influence of selection of model calibration approaches from a practical point of view. The case study is on calibration of the default Federal Highway Administration (FHWA) simplified facing load model for facing design of soil nail walls. A total of 23 measured facing load data collected by Liu et al. [15] are

adopted for calibration. Calibration results confirm that the two approaches are not practically equivalent when the data available for calibration are scarce. A model calibrated by the PDC approach usually does not reach minimal RMSE or vice versa. The Bayesian information criterion (BIC) is introduced to rank the competence of the PDC and RMSE-calibrated models fitting to the data. According to BIC, a model calibrated by the PDC approach may or may not be superior to its counterpart by the RMSE approach, depending on the assumption of distribution of model bias.

The two PDC- and RMSE-calibrated models are then used for estimation of facing loads and design of reinforcement ratio against the facing flexure limit state using both deterministic and reliability-based design approaches. It is demonstrated that the estimated facing loads and the determined reinforcement ratios using both calibrated models do not differ significantly from each other. Therefore, in practice, either approach can be adopted for geotechnical model calibration even with scarce data.

Last, there are also other approaches for model calibrations, for example, the Bayesian inference technique. The Bayesian approach provides distributions other than point estimates for estimation of model parameters. This sets the basic differences between the Bayesian approach and the

PDC and RMSE approaches. Discussion on parameter determination using Bayesian approaches can be referenced to, e.g., Lin and Yuan [39] and Lin et al. [40].

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful for financial support provided by Guangdong Provincial Key Laboratory of Marine Civil Engineering (LMCE202107).

References

- [1] Z. Cao, Y. Wang, and D. Li, *Probabilistic Approaches for Geotechnical Site Characterization and Slope Stability Analysis*, Springer, Verlag Berlin Heidelberg, 2017.
- [2] International Organization for Standardization, *General Principles on Reliability of Structures*, ISO2394:2015, Geneva, Switzerland, 2015.
- [3] K.-K. Phoon, "The story of statistics in geotechnical engineering," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 14, no. 1, pp. 1–23, 2019.
- [4] K.-K. Phoon, "The story of statistics in geotechnical engineering," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 14, no. 1, pp. 3–25, 2020.
- [5] K.-K. Phoon, J. Ching, and T. Shuku, "Challenges in data-driven site characterization," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pp. 1–13, 2021.
- [6] B. Bai, G.-c. Yang, T. Li, and G.-s. Yang, "A thermodynamic constitutive model with temperature effect based on particle rearrangement for geomaterials," *Mechanics of Materials*, vol. 139, Article ID 103180, 2019.
- [7] B. Bai, Q. Nie, Y. Zhang, X. Wang, and W. Hu, "Cotransport of heavy metals and SiO₂ particles at different temperatures by seepage," *Journal of Hydrology*, vol. 597, no. 2, Article ID 125771, 2020.
- [8] Y. Wu, J. Cui, J. Huang, W. Zhang, and N. Yoshimoto, "Correlation of critical state strength properties with particle shape and surface fractal dimension of clinker ash," *International Journal of Geomechanics*, vol. 21, no. 6, Article ID 04021071, 2021.
- [9] B. Yuan, M. Sun, L. Xiong, Q. Luo, and H. Li, "Investigation of 3D deformation of transparent soil around a laterally loaded pile based on a hydraulic gradient model test," *Journal of Building Engineering*, vol. 28, no. 6, Article ID 101024, 2019a.
- [10] B. Yuan, Z. Li, Z. Zhao, H. Ni, and Z. Li, "Experimental study of displacement field of layered soils surrounding laterally loaded pile based on transparent soil," *Journal of Soils and Sediments*, vol. 21, no. 4, pp. 1–12, 2021a.
- [11] B. Yuan, Z. Li, Z. Su, Q. Luo, M. Chen, and Z. Zhao, "Sensitivity of multistage fill slope based on finite element model," *Advances in Civil Engineering*, vol. 2021, Article ID 6622936, 13 pages, 2021b.
- [12] B. Yuan, Z. Li, Y. Chen et al., "Mechanical and microstructural properties of recycling granite residual soil reinforced with glass fiber and liquid-modified polyvinyl alcohol polymer - ScienceDirect," *Chemosphere*, vol. 286, Article ID 131652, 2021c.
- [13] L. Zhang, F. Wu, Y. Zheng, L. Chen, J. Zhang, and X. Li, "Probabilistic calibration of a coupled hydro-mechanical slope stability model with integration of multiple observations," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 12, no. 3, pp. 169–182, 2018.
- [14] P. Lin and J. Liu, "Evaluation and calibration of ultimate bond strength models for soil nails using maximum likelihood method," *Acta Geotechnica*, vol. 15, no. 7, pp. 1993–2015, 2020.
- [15] P. Lin, R. J. Bathurst, and J. Liu, "Statistical evaluation of the FHWA simplified method and modifications for predicting soil nail loads," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 143, no. 3, Article ID 04016107, 2017.
- [16] P. Lin, P. Ni, C. Guo, and G. Mei, "Mapping soil nail loads using Federal Highway Administration (FHWA) simplified models and artificial neural network technique," *Canadian Geotechnical Journal*, vol. 57, no. 6, pp. 1453–1471, 2020.
- [17] K.-K. Phoon and F. H. Kulhawy, "Characterisation of model uncertainties for laterally loaded rigid drilled shafts," *Geotechnique*, vol. 55, no. 1, pp. 45–54, 2005.
- [18] K.-K. Phoon and C. Tang, "Characterisation of geotechnical model uncertainty," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 13, no. 2, pp. 1–30, 2019.
- [19] C. Tang and K.-K. Phoon, "Statistics of model factors and consideration in reliability-based design of axially loaded helical piles," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 144, no. 8, Article ID 04018050, 2018.
- [20] J. Yuan, P. Lin, G. Mei, and Y. Hu, "Statistical prediction of deformations of soil nail walls," *Computers and Geotechnics*, vol. 115, Article ID 103168, 2019b.
- [21] J. Yuan, P. Lin, R. Huang, and Y. Que, "Statistical evaluation and calibration of two methods for predicting nail loads of soil nail walls in China," *Computers and Geotechnics*, vol. 108, pp. 269–279, 2019c.
- [22] R. J. Bathurst and Y. Yu, "Probabilistic prediction of reinforcement loads for steel MSE walls using a response surface method," *International Journal of Geomechanics*, vol. 18, no. 5, Article ID 04018027, 2018.
- [23] H. Liu, P. Lin, C. Guo, Z. Li, and X. Qin, "A shallow artificial neural network for mapping bond strength of soil nails," *Marine Georesources & Geotechnology*, vol. 39, no. 3, pp. 1–13, 2019.
- [24] Y. Yu and R. Bathurst, "Probabilistic assessment of reinforced soil wall performance using response surface method," *Geosynthetics International*, vol. 24, no. 5, pp. 524–542, 2017.
- [25] J. Zhang, H. Huang, and K. Phoon, "Application of the Kriging-based response surface method to the system reliability of soil slopes," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 139, no. 4, pp. 651–655, 2012.
- [26] J. Zhang, H. Chen, H. Huang, and Z. Luo, "Efficient response surface method for practical geotechnical reliability analysis," *Computers and Geotechnics*, vol. 69, pp. 496–505, 2015a.
- [27] W. Zhang, A. T. Goh, and F. Xuan, "A simple prediction model for wall deflection caused by braced excavation in clays," *Computers and Geotechnics*, vol. 63, pp. 67–72, 2015b.

- [28] W. Zhang, C. Wu, Y. Li, L. Wang, and P. Samui, "Assessment of pile drivability using random forest regression and multivariate adaptive regression splines," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 15, no. 1, pp. 1–14, 2019.
- [29] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [30] K. P. Burnham and D. R. Anderson, "Multimodel inference," *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004.
- [31] C. H. Juang, J. Ching, L. Wang, S. Khoshnevisan, and C.-S. Ku, "Simplified procedure for estimation of liquefaction-induced settlement and site-specific probabilistic settlement exceedance curve using cone penetration test (CPT)," *Canadian Geotechnical Journal*, vol. 50, no. 10, pp. 1055–1066, 2013.
- [32] H. Liu, L. Tang, P. Lin, and G. Mei, "Accuracy assessment of default and modified Federal Highway Administration (FHWA) simplified models for estimation of facing tensile forces of soil nail walls," *Canadian Geotechnical Journal*, vol. 55, no. 8, pp. 1104–1115, 2018.
- [33] C. Lazarte, H. Robinson, J. Gómez, A. Baxter, A. Cadden, and R. Berg, *Geotechnical Engineering Circular No. 7 Soil Nail walls—Reference Manual*, Rep. No. FHWA-NHI-14-007, Federal Highway Administration, Washington, DC, 2015.
- [34] P. Lin and R. J. Bathurst, "Influence of cross correlation between nominal load and resistance on reliability-based design for simple linear soil-structure limit states," *Canadian Geotechnical Journal*, vol. 55, no. 2, pp. 279–295, 2018a.
- [35] R. Byrne, D. Cotton, J. Porterfield, C. Wolschlag, and G. Ueblacker, *Manual for Design and Construction Monitoring of Soil Nail walls*, Federal Highway Administration, Washington, DC, 1998.
- [36] P. Lin and R. J. Bathurst, "Reliability-based internal limit state analysis and design of soil nails using different load and resistance models," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 144, no. 5, Article ID 04018022, 2018b.
- [37] P. Lin and R. J. Bathurst, "Calibration of resistance factors for load and resistance factor design of internal limit states of soil nail walls," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 145, no. 1, Article ID 04018100, 2019.
- [38] Aashto, *LRFD Bridge Design Specifications*, American Association of State Highway and Transportation Officials (AASHTO), Washington DC, 9th Ed. edition, 2020.
- [39] P. Lin and X.-X. Yuan, "A two-time-scale point process model of water main breaks for infrastructure asset management," *Water Research*, vol. 150, pp. 296–309, 2019.
- [40] P. Lin, X. X. Yuan, and E. Tovilla, "Integrative modeling of performance deterioration and maintenance effectiveness for infrastructure assets with missing condition data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 8, pp. 677–695, 2019.