Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

Comparative Study of the Effects of Three Data-Interpretation Methodologies on the Performance of Geotechnical Back Analysis.

WANG, Ze-Zhou[1,2]; Siang Huat, GOH[1,2]; Chan Ghee, KOH[1,2]; Ian F.C. SMITH[2,3]

1. Department of Civil & Environmental Engineering, National University of Singapore, Block E1A, #07-03, No.1 Engineering Drive 2, Singapore 117576. Email: e0054291@u.nus.edu (WANG,Ze-Zhou); gohsianghuat@nus.edu.sg (Siang Huat, GOH); cgkoh@nus.edu.sg (Chan Ghee, KOH)

2. Future Cities Laboratory, Singapore-ETH Centre, 1 CREATE Way, CREATE Tower, #06-01, Singapore 138602, Singapore

3. Applied Computing and Mechanics Laboratory (IMAC), School of Architecture, Civil and Environmental Engineering (ENAC), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. Email: ian.smith@epfl.ch (Ian F.C. SMITH)

Corresponding author:

WANG, Ze-Zhou

Email: e0054291@u.nus.edu

Postal address: Future Cities Laboratory, Singapore-ETH Centre, 1 Create Way, CREATE Tower, #06-01, Singapore 138602

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

ABSTRACT:

Back analysis can provide engineers with important information for better decision-making. Over the years, research on back analysis has focused mainly on optimisation techniques, while comparative studies of data interpretation methodologies have seldom been reported. This paper examines the use of three data-interpretation methodologies on the performance of geotechnical back analysis. In general, there are two types of approaches for interpreting model predictions using field measurements, deterministic vs population-based, both of which are considered in this study. The methodologies that are compared are (a) error-domain model falsification (EDMF), (b) Bayesian model updating and (c) residual minimisation. Back analyses of an excavation case history in Singapore using the three methodologies indicate that each has strengths and limitations. Residual minimisation, though easy to implement, shows limited capabilities of interpreting measurement data with large uncertainty errors. EDMF provides robustness against incomplete information of the correlation structure. This is achieved at the expense of precision, as EDMF yields wider confidence intervals of the identified parameter values and predicted quantities compared to Bayesian model updating. In this regard, a modified EDMF implementation is proposed which can improve upon the limitations of the traditional EDMF method, thus enhancing the quality of the identification outcomes.

KEYWORDS:

Excavation, Back analysis, Parameter identification, Bayesian updating, Observational method, Finite-element analysis

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

1. INTRODUCTION

Construction activities for underground structures are usually monitored. The rich information embedded in field-response measurements can be used to enhance knowledge of material parameter values and the overall behaviour of underground structures, leading to a potential reduction in construction risks and costs. The procedure whereby material parameter values are estimated from field-response measurements is often called back analysis, which is an essential step for implementing the observational method (Hardy et al[19]; Peck[32]).

Four components are needed for an effective implementation of back analyses. These include (a) a calculation model (b) field-response measurements (c) a data interpretation methodology and (d) an optimisation technique. A data-interpretation methodology is defined as the methodology that determines how model predictions are assessed and interpreted given field-response measurements. An optimisation algorithm is used to facilitate the search for solutions based on the interpretations specified by the data-interpretation methodology.

Over the years, research into geotechnical back analysis has focused mainly on optimisation techniques. The performance of optimisation techniques, such as gradient-based techniques (Finno and Calvello[12]), Heuristic algorithms (Levasseur et al[27]; Knabe et al[24]), surrogate-based optimisation (Qi and Zhou[36]; Pai et al[35]; Zhang et al[45]) and multi-objective optimisation (Huang et al[18]; Jin et al[23]), has been studied in the context of geotechnical back analysis. A comparative study on the performance of these optimisation algorithms has also been reported (Yin et al[44]). Despite receiving less attention to date, other data-interpretation methodologies, such as probabilistic approaches, play an important role because they determine, in the presence of uncertainty, how model predictions are assessed and interpreted given field-response measurements. Only when model predictions are interpreted in a reliable manner can the benefits of advanced optimisation techniques be fully realised. There are in general two approaches for interpreting field measurements using model predictions, deterministic and

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

population-based approaches. Residual-minimisation (Finno and Calvello[12]; Jofre[22]; Ledesma et al[25]) is a deterministic approach that is commonly reported in the literature, while Bayesian model updating (Juang et al[17]; Qi and Zhou[36]) is the most commonly adopted population-based approach. More recently, the population-based error-domain model falsification (EDMF) methodology, which was previously used for bridge engineering (Cao et al[9]; Goulet et al[15]; Proverbio et al[34]), leakage detection (Moser et al[30]) and wind engineering (Vernay et al[42]), has also been applied and adapted for geotechnical excavation back analysis (Wang et al[43]). To date, however, these methodologies have not been systematically compared for the same geotechnical back analysis case.

In this paper, a comparative study of methodologies involving residual minimisation, Bayesian model updating and error-domain model falsification is presented. Field measurements from an excavation case history in Singapore are used to provide information for data interpretation. Results are presented and discussed to highlight the strengths and limitations of each methodology. Arising from this study, a modified EDMF implementation is also proposed, and improved performance is illustrated through a comparison with the results from the traditional EDMF implementation.

2. DATA-INTERPRETATION METHODOLOGIES

2.1. Error-domain model falsification (EDMF)

This method was developed based on the assertion of Sir Karl Popper in *The Logic of Scientific Discovery* (Popper[33]) that models cannot be fully validated by data and that they can only be falsified. In the context of EDMF, the analysis starts from an initial population of material parameter sets. The falsification process is then carried out to eliminate parameter sets that do not yield predictions compatible with field-response measurements, based on some pre-defined acceptance criteria. The remaining non-falsified parameter sets, which are termed candidates,

Accepted Manuscript

are considered as viable inputs for use with the numerical model to assess the behaviour of the actual system. In this regard, EDMF often yields a solution that comprises a population of candidates. The more detailed mathematical formulation is provided below.

A plausible physics-based model defined by n_θ parameter values and a model class G_k can be identified using information provided by field-responses measurements. In the context of an excavation problem, such measurements may take the form of a retaining wall deflection profile obtained from inclinometer readings taken at n_v number of measurement locations. Let R_i and $\hat{y}_I$ denote the real response and the measured response respectively at location i $\in\{1, \ldots, n_v\}$. Values for Θ'_k, which correspond to the true parameter values, can be assigned to obtain predictions $g_{i,k}(\Theta'_k)$ of the model class at location i. Modelling uncertainties arising from model simplifications and omissions and measurement uncertainties are expressed as $U_{i,gk}$ and $U_{i,\hat{y}}$ respectively at location i. The mathematical relationship between these quantities is given in Equation 1:

$$g_{i,k}(\Theta'_k)+ U_{i,g_k} = R_i = \hat{y}_i + U_{i,\hat{y}} \quad \forall_i \in\{1,\ldots,n_y\} \tag{1}$$

Upon rearrangement, the Equation 2 is obtained:

$$g_{i,k}(\Theta'_k)-\hat{y}_i=U_{i,ck} \tag{2}$$

where $U_{i,ck}$ is a random variable representing the difference between the measurement uncertainty $U_{i,\hat{y}}$ and the modelling uncertainty $U_{i,gk}$ at location i.

The left term of Equation 2 represents the difference between model prediction and measurement data at location i. This term is typically called the residual r_i. The probability

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

density function (PDF) describing the modelling uncertainty in the model class $f_{Ui,gk}(u_{i,gk})$ can be estimated and applied in the analysis.

The implementation of EDMF starts with the definition of an initial model set, which contains n_Ω model instances $\Omega_k = \{\Theta_{k,m},\ m = 1,\ldots,\ n_\Omega\}$. Threshold bounds are then defined by computing the narrowest interval $\{u_{i,low},\ u_{i,high}\}$ that represents a probability equal to $\varnothing_d^{1/n_v}$ for the combined PDFs $f_{Ui,c}(u_{i,c})$ at each measurement location i. This computation is performed using the following equation:

$$\varnothing_d^{1/n_v} = \int_{u_{i,low}}^{u_{i,high}} f_{u_{i,c}}\left(u_{i,c}\right) du_{i,c} \quad \forall_i \in \{1,\ldots,n_v\} \tag{3}$$

A value of 0.95 for the confidence level $\varnothing_d \in [0,1]$ is commonly employed. The confidence level $\varnothing_d$ is adjusted using the Šidák correction (Abdi[1]; Šidák[40]) to take into account the fact that n_v measurement locations are simultaneously considered. Uniform probability distributions create a hyper-rectangular acceptance region. Under this scheme, the correlation information between sensor locations is no longer needed, which is particularly helpful because it is often difficult to determine the correlation values between sensor locations. Goulet et al[15] have shown that this scheme is conservative in many situations. Falsification is then performed according to the following equation:

$$\Omega_k'' = \{\Theta_k \in \Omega_k \mid \forall_i \in \{1,\ldots,n_v\}\ u_{i,low} \le g_{i,k}(\Theta_k) - \hat{y}_i \le u_{i,high}\} \tag{4}$$

where the candidate model set (CMS), Ω_k'', is made up of all model instances that have not been falsified. An instance Θ^*_k of a model class G_k is a candidate model if for each sensor location $i \in \{1,\ \ldots,\ n_v\}$, the residual r_i value falls inside the threshold bounds derived from Equation 3.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

All model instances that belong to the CMS are assigned a constant probability, as it is often difficult to justify a more sophisticated distribution in practical situations.

In large-scale, multi-staged excavations, a continuous falsification and prediction framework can be established. At an early or intermediate point in the excavation process, it is often useful and desirable to perform back analysis with field-response measurements and predict the field responses of subsequent excavation stages with the identified material parameter values. In the context of EDMF, predictions are made with all candidate models. Detailed mathematical development of EDMF can be found in Goulet and Smith[16] and the implementation of it on a multi-stage excavation problem can be found in Wang et al[43].

2.2. Bayesian model updating

Bayesian model updating starts from the evaluation of the likelihood function, which is defined as the joint probability that the residuals r are equal to the values computed with a given set of material parameter value, modelling uncertainty and measurement uncertainty. Following the notation in Section 2.1, the likelihood function can be expressed in Equation 5:

$$p(r;\mu_{U_{ck}},C)=\frac{1}{2\pi^{\frac{n_v}{2}}|C|^{\frac{1}{2}}}\exp^{\{-\frac{1}{2}\left(r-\mu_{U_{ck}}\right)^T C^{-1}\left(r-\mu_{U_{ck}}\right)\}} \quad (5)$$

where r denotes the difference between model predictions and measurement data. Equation 5 is formulated based on the case wherein observations of multiple points are available and the modelling and the measurement uncertainties are normally distributed. Therefore, the combined uncertainty U_{ck} also follows a normal distribution with mean $\mu_{U_{ck}}$, which is the sum of the mean of all modelling and measurement uncertainties, and standard deviation $\sigma_{U_{ck}}$, which is the square root of the sum of the variance of all uncertainties. The joint probability of

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

residuals at multiple measurement locations can then be calculated with a chosen correlation matrix C, which describes the dependency of uncertainties at these measurement locations. Two correlation schemes are examined in this study. The first scheme assumes independence or zero correlation. This scheme suggests that the magnitudes of the uncertainties at measurement locations are independent from each other. The second scheme adopts a correlation matrix following the work of Qi and Zhou[36] and Ledesma et al[25]. The effects of correlation matrix will be discussed later in this paper.

The correlation structure C is an input for the evaluation of the likelihood, and has to be estimated before the posterior distribution can be obtained. In Bayesian methodology, it is assumed that the correlations are independent from the magnitude of the uncertainties. According to Qi and Zhou[36] and Ledesma et al[25], the correlation structure may be calculated as follows:

$$C_{ij} = \sigma^2 \sum_{m=1}^{\min(i,j)} l_m^2 \tag{6}$$

where

C_{ij} is the covariance of the measurement uncertainties at point i and j,

σ is the standard deviation of the inclinometer, which can be estimated following the work of Finno and Calvello[12], and

l_m is the distance between two neighboring measurement points, which is 1m in the current study.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

In the Bayesian framework, the posterior distribution, which describes the conditional probability of the material parameter values given the field-response measurements, can be calculated using Equation 7:

$$p(\Omega_k|\hat{y}_i) = k_1 \cdot p(r;\mu_{U_{ck}},C) \cdot p(\Omega_k) \quad (7)$$

where k_1 is a normalization constant that makes the probability density function valid, and $p(\Omega_k)$ represents the prior probability density function of the material parameter values to be identified. In the current study, the prior probability density function is assumed to be uniformly distributed. Also, both the modelling and measurement uncertainties are explicitly included in the evaluation of the likelihood function, which provides a basis to compare results from the Bayesian model updating approach with those obtained from EDMF analysis. In this paper, the implementation of the Bayesian model updating calculations is facilitated by the use of the Markov-chain Monte-Carlo simulation and the Metropolis-Hastings algorithm, in conjunction with the response surface method. Details of the implementation can be found in Juang et al[17] and Qi and Zhou[36].

2.3. Residual minimization

The residual minimization approach usually follows the weighted least-squares criterion, which aims to find a single set of parameter values that produces the minimum absolute error between numerical predictions and measurements. Finno and Calvello[12] reported its application to an excavation back analysis problem. The objective function is expressed as:

$$S(\Omega_k)=[g_{i,k}(\Omega_k) - \hat{y}_i]^T \omega \left[g_{i,k}(\Omega_k) - \hat{y}_i \right] = r_i^T \omega r_i \quad (8)$$

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

where ω is the diagonal weighting matrix whose values correspond to the inverse of the measurement error variances. It is noted that the implementation reported by Finno and Calvello[12] does not include errors associated with modelling uncertainties.

In addition to measurement errors, the residual minimisation implementation in this study also accounts for errors due to modelling uncertainties. Both measurement and modelling uncertainties are assumed to be normally distributed, and their respective mean values are adopted as the measurement and modelling errors for the deterministic calculations. The modified objective function is expressed in Equation 9:

$$S(\Omega_k)=[g_{i,k}(\Omega_k) - \mu_{U_{i,ck}} - \hat{y}_i]^T \omega \left[g_{i,k}(\Omega_k) - \mu_{U_{i,ck}} - \hat{y}_i \right] = r_i^T \omega r_i \quad (9)$$

Such a formulation allows a non-zero mean estimate of the modelling uncertainties to be included in the analysis, which is consistent with the consideration of such modelling errors in both the Bayesian model updating and the EDMF approaches.

The search for the optimal solution is facilitated by the response surface method, which will be explained later. The search starts with a dense grid of parameter values, from which several local minima may be identified. It is noted that the search-based objective function associated with the absolute minimum derived from the original grid may not necessarily correspond to the global minimum. For this reason, refined searches are performed in the vicinities of the various local minimum zones identified. Such refinements around the local minima are carried out to account for the possibility that the parameter combination that gave the best objective function value may be located near a local minimum of the initial grid. The refinement is repeated for each local minimum three times. The final parameter combination is taken as the

Accepted Manuscript

one that gave the best objective function values across the refined grids of all the local minima. A similar procedure has been reported by Ieronymaki et al[20, 21].

3. CASE STUDY

Back-analyses of an excavation case history in Singapore are performed using the three data-interpretation methodologies described in Section 2. The geotechnical finite-element software package Plaxis 2D (Brinkgreve et al[8]) and Plaxis 3D (Brinkgreve et al[7]) are used to model the excavation and the wall deflection response. Due to the large number of simulations required, the response surface method is adopted to facilitate the search of solutions. Three-dimensional excavation effects arising from the corner constraints are quantified using the approach proposed by Wang et al[43].

3.1. Project description

The excavation is approximately 60m in length, 40m in width and 10m in depth. The support system of the excavation includes diaphragm walls, soldier pile walls, toe pins and two layers of steel struts and waler beams. Figure 1 shows the 3D finite-element model. The 800mm thick diaphragm walls are modelled as elastic plate members. The reduced lateral stiffness of the diaphragm walls due to the presence of construction joints (Dong et al[10]; Zdravkovic et al[46]) between panels is captured by releasing the rotational stiffness between the plates, following Lee et al[28]. Along certain sections of the excavation, the top 1 to 3m of the diaphragm walls are replaced by soldier pile walls, comprising grade 355 steel universal column sections and 75mm thick concrete panels. The toe pins, which extend 2m below the toe of the diaphragm walls, consist of grade 355 circular steel hollow sections placed at 800mm spacing. As a simplification, the toe pins and the soldier pile walls are smeared and modelled as elastic plate members with equivalent properties. Struts and waler beams are modelled as node-to-node

Accepted Manuscript

anchors and beam elements respectively. Interface elements with zero thickness (Brinkgreve et al[7, 8]) are used to model the soil-wall interactions. The properties of all structural elements are listed in Table 1.

The soil stratigraphy in the 3D model was based on the information obtained from six boreholes drilled at this site, which is situated on the Bukit Timah Granite formation. The top layer, which is roughly 3m in thickness, contains mostly sandy silt and man-made backfill materials. It is underlain by a 10 to 13m thick residual soil layer of sandy silt, denoted as G(VI), across most parts of the project site. The granitic rock layer G(III) is present at approximately 15m below the ground surface. On the eastern half of the project site, there is also a 5m thick layer of coarse sand sandwiched between the G(VI) sandy silt and G(III) granitic rock. In addition, the SI report indicates that a pocket of medium to coarse gravels is present at a localised area near the centre of the pit.

Figure 2 shows the 2D finite-element model of the west-to-east middle section of the excavation. The idealized 2D soil profile, excavation support system and boundary conditions are also shown in the figure. Roller supports are assigned to vertical boundaries while the base of the model is fully fixed. The fill layer and the gravels are described using the Mohr-Coulomb model while the rock layer is described using the Hoek-Brown model. Other soil layers are simulated using the Hardening Soil with Small Strain Stiffness (HS Small) model (Benz[6]). Representative soil parameters for the Bukit Timah formation (Rahardjo et al[37, 39]) are adopted and listed in Table 2, except for the grey shaded cells which indicate the parameters to be identified.

In this study, four parameters are selected to be identified. They are (a) the Young's modulus E (MPa) of the fill layer, (b) the reference Young's modulus E_{50}^{ref}(MPa) of the G(VI) sandy silt layer, (c) the reference Young's modulus E_{50}^{ref}(MPa) of the coarse sand layer and (d) the equivalent flexural rigidity EI (kNm^2) of the smeared soldier pile walls. These four parameters

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

are selected based on the results of a sensitivity analysis. Preliminary estimated ranges of these parameter values at the start of the identification process are indicated in the shaded cells of Tables 1 and 2. Other Hardening Soil model reference moduli, e.g. E_{oed}^{ref}, E_{ur}^{ref} and G_0^{ref}, of the sandy silt and coarse sand layers are correlated to E_{50}^{ref}, as shown in the tables. The initial water table is 2m below the ground level.

The construction sequence modelled in the finite-element analysis comprises 6 stages, as shown on Table 3. As the soil layers are inclined with varying thicknesses, the initial ground stresses in stage 0A are generated using the gravity turn-on approach. The diaphragm wall is 'wished-in-place' in stage 0B, assuming negligible installation effects. Fully coupled flow-deformation calculations (Galavi[14]) are performed to account for the combined time-dependent effects arising from groundwater seepage and consolidation. As indicated in Figures 1 and 2, wall deflection measurements of inclinometers 04 and 09, located at the east and west sides of the excavation respectively, are included in the back analysis.

3.2. Response surface method

Back analysis involves repeated evaluations of the finite-element model using different combinations of the material parameters to be identified. Depending on the number of parameters and the initial ranges adopted, the number of combinations, and hence finite-element simulations, may run into the thousands, which is impractical from a computational point of view. To reduce the computational time to a manageable level, the response surface method may be used as a surrogate for the finite-element analysis. In this method, a smaller but adequate number of 2-D finite element analyses are first performed to obtain mathematical functions that can reasonably relate the material parameters of interest (e.g. Young's modulus E of the fill layer) to the field quantities of interest (e.g. wall deflection at a given depth). Such mathematical functions are then used in the back-analysis to obtain predictions of the field

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

quantities of interest for several thousand combinations of the material parameters, which can tremendously reduce the computational effort by obviating the need to perform finite element analyses for all these cases.

In this study, the three-dimensional excavation effects are quantified following the approach of Wang et al.[43]. In this method, the 2D finite element model and its associated response surfaces are used to perform the bulk computations of the back analysis for parameter identification. The construction of the response surface starts from the generation of initial sampling points. In the current study, initial sampling points are generated by combining (i) the central composite design (Ahmadi et al[2]), which generates 36 samples of material parameter combinations, and (ii) Latin Hypercube sampling technique (Stein[41]), which generates an additional 100 combination samples. In total, these provide 136 parameter combinations for which 2D finite element analyses are performed and the results used to construct the response surfaces. The results of the 136 finite-element analyses are mathematically related to the four parameters to be identified using the Gaussian process regression model (Rasmussen[38]), which was found to perform better than the polynomial regression model and radial-basis function method for this study. The machine learning toolbox in Matlab was used to generate the regression models with a quadratic basis function and an exponential kernel function, which performed better than other settings available in Matlab.

As the back analyses involve wall deflection measurements made at different excavation stages and multiple measurement locations/depths, it is necessary to generate multiple response surfaces, each corresponding to a particular excavation stage and measurement location/depth. In this study, the wall deflections measured by inclinometers 04 and 09 are used, and their locations are indicated in Figures 1 and 2. A total of 106 response surfaces are needed to cover all measurement points of the four excavation stages. These response surfaces are then

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

validated using an additional 50 combination samples of the four parameter values that are not used in the construction of the response surfaces.

3.3. Quantification of the three-dimensional effects of excavation

As shown in Figures 1 and 2, the box-shaped geometry of the excavation are expected to introduce corner constraints, the effects of which cannot be captured using the 2D model (Ou et al[31]). While 3D analyses can be performed to generate response surfaces that account for such corner effects, the large number of 3D simulations required makes this an impractical option. The method proposed by Wang et al.[43] quantifies corner effects approximately using 2D analysis by introducing an error term to the wall deflection obtained from the plane-strain-based computations. At any particular measurement depth and excavation stage, the specific error term is calculated as the difference between the 2D and the 3D finite-element wall deflections (= $2D_{FE}$ wall deflection – $3D_{FE}$ wall deflection). Wang et al.[43] showed that, using this method, it is possible to approximately quantify the three-dimensional effects at all measurement locations and excavation stages by performing only two 3D finite element analyses.

4. COMPARATIVE STUDY

A sound back analysis should (a) identify reasonably accurate parameter values and (b) provide good predictions of the quantities of interest. In the subsequent sections of the paper, the performance of all three data-interpretation methodologies is independently assessed from these two perspectives.

During underground construction, it is often useful and desirable to predict the field responses of later excavation stages using measurements from the early or intermediate excavation stages. In the current study, four major excavation stages are considered. Hence, a back analysis can

Accepted Manuscript

be performed after each major excavation stage. In the subsequent discussions, the term '1^{st} round of identification' refers to the back analysis performed after excavation stage 1, using only wall deflection measurements of excavation stage 1. Similarly, the term '4^{th} round of identification' refers to the back analysis performed after excavation stage 4, wherein wall deflection measurements of excavation stages 1 to 4 are utilised.

All three methodologies adopt the same set of modelling and measurement uncertainties. These uncertainty sources include inclinometer errors, model simplification from 3D to 2D, errors arising from the use of response surfaces, and others. The magnitudes of these uncertainties are obtained from the values reported in the literature (Goulet et al[15]; Finno and Calvelo[12]) and quantification methods proposed (Wang et al[43]). Examples of typical uncertainty ranges computed based on Equation 3 are shown in Table 4.

4.1. Results of Bayesian model updating

In the current study, 5000 Markov chain samples were simulated. The scaling factors adopted in the Markov chain simulations were determined based on a trial-and-error approach, which was guided by checking the means and the standard deviations of the posterior distributions as a function of the number of Markov chain samples. The scaling factors determined are reasonable, with the means and the standard deviations of the posterior distributions converging to stable values within the 5000 Markov chain samples simulated.

As discussed in Section 2.2, the effects of the correlation matrix on the performance of the Bayesian approach will be examined. Two correlation schemes are implemented in this study, one assuming independence (or zero correlation) while the other follows the proposed correlation of Qi and Zhou[36] and Ledesma et al[25]. The non-zero matrix of correlation coefficients at selected measurement points are shown in Table 5.

Accepted Manuscript

Figure 3 shows posterior distributions of the four identified parameters obtained via Bayesian analysis using measurement data from all four excavation stages. It is noted that, apart from the E_{50}^{ref} value of the silt layer shown in Figure 3b, the posterior distributions differ quite significantly for the two correlation schemes. The mean values of E of the fill layer, E_{50}^{ref} of the sand layer and EI of the soldier pile wall differ by approximately 90%, 75% and 40% respectively across the two correlation schemes. Figure 3 also shows the typical parameter values estimated using empirical correlations with blow count data (N values) from standard penetration tests (SPT). Two typical correlations, 1.5 times the SPT-N value and 2 times the SPT-N value, are considered. Comparing the results of the Bayesian analysis with the empirically estimated parameter values, the zero correlation assumption appears to yield values that are better in line with the empirical estimates. In addition, parameter values identified with the zero correlation scheme are also more consistent and in line with laboratory and field test values reported by Leung et al[26]; Moon et al[29]; Zhang et al[47] on similar soil types.

Figure 4 shows, for the case of non-zero correlation, the mean and the 95% bounds of the predicted wall deflection profiles of ID 09 and ID 04 at the final excavation stage, after each round of identification. The bounds are calculated considering both the modelling and the measurement uncertainties. From the upper row of subplots for wall ID 09, it is seen that the mean predictions of wall ID 09 for each stage can capture the bulging wall profile quite reasonably. While the first three rounds of identification tend to under-predict the maximum wall deflection at the final excavation stage by about 20 to 30%, the 4th round of identification leads to a good overall prediction with the maximum deflection under-estimated by about 10%. Although the mean predictions tend to under-estimate the measurements, the 95% bounds of the predicted wall deflection profiles reasonably enclose the measurements.

In contrast, the lower row subplots of Figure 4 show that the agreement between the predicted and measured deflection responses of wall ID 04 are not as good. This could be due in part to

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

the 'composite' wall section at ID 04, which comprises a flexible soldier pile wall for the top 3m and a much stiffer diaphragm wall below, as delineated by the horizontal line at Reduced Level 121 in subplots (e) to (h) of Figure 4. The presence of the more flexible soldier pile wall results in larger measured deflections near the top of wall ID 04, so that the profile exhibits two peaks, instead of the more typical bulging profile of ID 09. However, the Bayesian predictions for ID 04 do not exhibit such a double peak profile. Furthermore, for both ID 04 and ID 09, the upper bounds of the predicted peak deflections fall very close to the measurement data in the 2nd and 3rd round of identification. This trend persists for ID 04 even after the 4th round of identification, despite incorporating measurement data from the final excavation stage itself. While the Bayesian model updating results are generally reasonable in that the measured deflections of IDs 04 and 09 as a whole fall within the two computed bounds, the proximity of the ID 04 maximum measured deflections to the predicted upper bound suggests that the predictions of ID 04 are less effectively updated as compared to those of ID 09.

An examination of the likelihood values computed by the Bayesian analysis reveal a possible reason behind the observed issues related to wall ID 04 highlighted in the previous paragraph. The process of Bayesian model updating involves the calculation of the likelihood function, which is defined as the joint probability of the measurements with a given set of material parameter values, modelling uncertainty and measurement uncertainty. In the current example, measurement data from two separate inclinometers (ID 04 and ID 09) are lumped together and considered simultaneously in the analysis, which resulted in some measurement points from ID 09 exerting a stronger influence in the calculation of the joint likelihood. A total of 106 measurement points are considered in the 4th round of identification. Among these 106 points, only 11 measurement points are associated with wall ID 04 at the final excavation stage. It turned out that the calculated joint likelihood from the Bayesian analysis does not differ much with and without the consideration of these 11 points, which suggest that they have weak

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

influence on the model updating calculations. In contrast, the differences in the joint likelihood computed with and without consideration of the final excavation measurements of wall ID 09 are more significant.

An additional Bayesian analysis is performed for the same excavation, but this time using only the measurements of wall ID 04. Figure 5 shows the comparison of the computed and measured wall deflections for walls ID 04 and ID 09 after the 4^{th} round of identification. Despite using only measurement data of wall ID 04, the predicted mean wall deflections of ID 04 and ID 09 agree reasonably well with the measured profiles. In fact, the improvement is quite significant for wall ID 04, in which the overall shape and magnitude of the deflection profile is quite well captured (compared to Figure 4) by the mean predictions. Figure 5 also shows that the Bayesian analysis is able to provide reasonable predictions of both the bounds and mean deflections at wall ID 09, even though the measurement data at this location was not considered in the model updating calculations.

4.2. Results of EDMF

The mathematical formulation of EDMF seeks to address the correlation-related issues encountered in Bayesian model updating. As explained in Section 2.1, Šidák correction and the rectangular acceptance region are implemented in EDMF analyses to circumvent the need for correlations. The use of such techniques produces exact results under independent or uncorrelated conditions, and conservative results when correlations exist (Farcomeni[13]; Goulet et al[15]; Šidák[40]). In the current study, 20,000 initial model instances corresponding to different parameter combinations are generated, and the associated wall predictions are made using the response surfaces generated from the 136 2D-finite element simulations discussed in Section 3.2. EDMF analyses are then performed to identify candidate models from these initial model instances.

Accepted Manuscript

Figure 6 shows the parallel-axis plot of the identified parameter values after the 4^{th} round of identification. Such a plot provides a way to visualise the values and trends of the identified parameters. The horizontal axis contains, at discrete spacing, four vertical axes corresponding to the four parameters to be identified. Each vertical axis plots the values of the corresponding parameter which it represents. A parameter combination set is represented by a line that connects the parameter values across the four vertical axes. Figure 6 contains 20,000 grey lines representing the 20,000 initial model instances, with the red dashed lines representing the 678 candidate models identified after the 4^{th} round of identification.

As plotted on the 2^{nd} vertical axis from the left, the identified E_{50}^{ref} values of the G(VI) silt layer fall into two clusters. The first cluster contains parameter values ranging from about 14MPa to 30MPa, while the second cluster comprises a smaller range of values between 47MPa and 50MPa. There is thus a conspicuous gap in the E_{50}^{ref} values between 30MPa and 47MPa. To check if this gap is due to inadequate sampling, the E_{50}^{ref} values between 30MPa and 45MPa are sampled at smaller intervals for the EDMF analysis. Even with such sampling refinements, the gap in the identified E_{50}^{ref} values still persists, indicating that its presence is not caused by inadequate sampling. Among the two clusters, the identified E_{50}^{ref} values of about 47 to 50MPa contained within the second cluster appear to be too high based on the average measured SPT-N value of about 18 for this G(VI) silt layer. Furthermore, parameter values reported in the literature (Leung et al[26]; Moon et al[29]; Wang et al[43]; Zhang et al[47]) suggest that this parameter should not exceed 35MPa. The identified silt layer E_{50}^{ref} values of 47 to 50MPa in the second cluster are thus likely to be erroneous. As discussed below, the parameter values in the second cluster are examples of Type II errors in statistical hypothesis testing.

In the current study, measurement data at multiple depths along the wall are utilized for parameter identification, the process of which involves performing multiple hypothesis testing. Table 6 summarizes all possible outcomes of a multiple hypothesis test. While the null

Accepted Manuscript

hypothesis can be either true or false, the outcome of the hypothesis testing can indicate the null hypothesis to be false even though it is actually true. When this happens, a Type I error is committed, which is denoted as $N_{1|0}$ in Table 6. Conversely, the outcome of the hypothesis testing can indicate the null hypothesis to be true even though it is actually false. This results in Type II errors, which is denoted as $N_{0|1}$ in Table 6. In this respect, the Šidák correction conservatively limits the probability of committing a Type I error by reducing the significance level of each individual test (Farcomeni[13]). However, type II errors are more likely to be committed. To improve the quality of predictions, it is important to minimize the population of Type II errors. In this study, an improved EDMF implementation that aims to reduce the number of Type II errors being committed is proposed and described in Section 5.

Figure 7 shows the parameter values identified by EDMF analysis, together with the posterior distributions obtained by Bayesian model updating. In general, the ranges of parameter values identified with EDMF are larger than the corresponding posterior distributions identified with the Bayesian approach. In fact, the EDMF analysis identifies parameter bounds that span the two posterior distributions calculated by the Bayesian analysis. This is not unreasonable given that the EDMF approach adopts a rectangular acceptance region, which implicitly allows all possible correlation configurations.

Figure 8 shows the the mean and the 95% bounds of predicted wall deflections at the final excavation stage after each round of identification. Compared to the predicted wall deflections shown in Figure 4 using the Bayesian technique, the EDMF analyses produce improved agreement between the predicted and measured wall deflections, especially for wall ID 04. This is reflected in the EDMF predicted mean deflection profile of ID 04, which shows reasonably good agreement with the measurement data. This is in contrast to the trend shown in Figure 4, wherein the measured maximum deflections fall closer to the predicted upper bound from the Bayesian analysis. The EDMF methodology requiring that the falsification check be

Accepted Manuscript

individually performed on each and every measurement point ensures that the measurement data at all measurement locations are accorded equal weightage. This is different from the Bayesian approach which evaluates a joint likelihood value for all measurement data. Such a function may be insensitive to the contributions of certain sets of measurement data, as was noted in Section 4.1 in which the effect of the ID 04 data was overshadowed by that of ID 09 when both sets of measurements were considered simultaneously. The comparison of predictions and bounds shown in Figure 4 and 8 indicate that EDMF produced improved results when measurement data of two inclinometers at different locations are simultaneously utilised for interpretation.

Figures 8e, 8f and 8g show that, after the first three rounds of identification, the mean predictions of the ID 04 wall deflections for the final excavation stage exhibit wall movements near the ground level that are larger than the measured values. These discrepancies indicate that the identified material parameters after the 3rd round of identification, which are based only on wall measurements taken from stages 1 to 3, are not sufficiently accurate to produce good predictions of the wall behaviour at this zone. With the additional deflection measurements from the final excavation stage included in the analysis, Figure 8h shows that the predictions near the top of wall ID 04 are improved.

In addition, the slightly wider bounds in Figure 8 as compared to those in Figure 4 indicate that EDMF predicts slightly larger variations in wall deflections. This is consistent with the larger variations in parameter values among the EDMF candidate models compared with the posterior distributions of Bayesian analyses, as shown in Figure 7.

4.3. Results of the residual minimisation approach

Using the residual minimization approach, the parameter values identified after 4th round of identification are shown by the dotted lines in Figure 9. Being a deterministic approach, the

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

residual minimisation method identifies a specific and unique combination of material parameter values, and therefore information on the variations in the parameter values are not available. Figure 9 shows that the E_{50}^{ref} values of silt layers and E_{50}^{ref} values of sand layers are reasonably compatible across the three data-interpretation methodologies. However, there are differences in the E values of fill layer and EI values of soldier pile wall. Such differences could be caused by inclinometer errors, which accumulate upward from the toe of the wall to the ground level. Deflection measurements taken near to the fill layer and the upper soldier pile wall are therefore subjected to larger measurement uncertainties, which, according to Equation 9, will be weighted less than the deeper measurements, which have smaller errors. Therefore, the E values of the fill layer and EI values of the soldier pile wall, which are both located in the shallower zones where measurement errors are likely higher, may not be effectively identified. In contrast, population-based approaches such as EDMF and Bayesian model updating can account explicitly for the mean and standard deviation of the uncertainties, and hence are better able to identify reasonable ranges of the material parameters.

Figure 10 shows the predicted wall deflections at the final stage of excavation after each round of identification using the residual minimization approach. After the 3^{rd} and 4^{th} round of identification, the predicted deflection profiles of wall ID 09 agree quite well with measurements, and are also quite similar to the Bayesian mean predicted deflections shown in Figure 4. However, the residual minimization predictions for wall ID 04 are not as good, especially near the top where the soldier pile wall is present. This observation is similar to the Bayesian results shown in Figure 4. As the residual minimization objective function evaluates the sum of the squared residuals of all measurement points, it is also likely that some data, in this case the 11 measurements of wall ID 04 at the final excavation stage, exerts a much smaller influence on the objective function compared to ID 09. This accounts for the discrepancies between the predictions and measurements for wall ID 04.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

5. A MODIFIED EDMF IMPLEMENTATION

A modified EDMF implementation is proposed in this section. It addresses the issue encountered in Section 4.2, where two clusters of E_{50}^{ref} values are identified for the silt layer using the conventional EDMF approach. The higher E_{50}^{ref} values contained in the smaller cluster are postulated as being Type II errors arising from the use of the Šidák correction technique when performing multiple hypothesis testing.

The work of De et al.[11] is perhaps the first to propose the use of the Benjamini-Hochberg (BH) correction technique (Benjamini and Hochberg[3]) as an alternative to Šidák correction in EDMF analysis. However, their validations were performed on conceptual problems under idealized problem settings and known values of the parameters to be identified. In contrast, full-scale engineering challenges are seldom so well-defined. In the following section, both traditional Benjamini-Hochberg correction (Benjamini and Hochberg[3]) and adaptive Benjamini-Hochberg correction (Benjamini and Hochberg[4]; Benjamini et al.[5]) will be implemented for the EDMF analysis of the present excavation case study. A comparison of identification results with these three correction techniques are also presented.

5.1. Challenge of multiple measurements

The Šidák correction, which is conventionally adopted in the EDMF implementation, keeps the probability of committing a Type I error at an acceptable level (e.g. 5%) by adjusting the bounds of the individual test when there is more than one measurement. The increasing number of measurements reduces the significance level of the individual tests, and hence the bounds of the individual tests are pushed farther back along the tails of the null hypothesis distribution. While this maintains the overall reliability of identification at 5%, it may also result in the occurrence of more Type II errors. To address this limitation, Benjamini and Hochberg[3, 4] proposed a correction technique that restrains the false discovery rate, defined as the expected

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

ratio of erroneous rejections to the number of rejected hypotheses. The use of the BH correction technique results in an improved restraints on the Type II errors, without adversely affecting the avoidance of a Type I error. In other words, the BH correction allows for a stronger rejection power than the Šidák correction and hence, its use will result in a smaller candidate model set because fewer Type II errors are committed.

5.2. Adaptive Benjamini-Hochberg correction

In contrast with the traditional Benjamini-Hochberg correction which adopts a fixed correction procedure, the adaptive Benjamini-Hochberg correction involves the estimation of the proportion of the null hypotheses that are actually true. These hypotheses are labelled as M_0. The estimation can be visualized through a graphical approach. The procedures are summarised by Benjamini and Hochberg[4] as follows:

a) Calculate the p-values, which defines the probability that the null hypothesis is true, of all residuals given the combined uncertainties, and arrange the p-values from smallest to the largest. In this study, they are annotated as p_i, $p_{(i+1)}, \ldots p_{(m)}$.

b) Calculate $S_i = (1- p_i)/(m+1-i)$, the i-th slope estimate.

c) Starting with $i = 1$, loop through $i = i + 1$ as long as $S_i \geq S_{i-1}$; stop at the first occurrence of $S_j \geq S_{j-1}$, and evaluate $\widehat{m}_0=\min[(1/S_j +1), m]$.

d) Compare each $p_{(i)}$ to $0.05i/\widehat{m}_0$ and reject the p-values for which $p_{(i)} \leq 0.05i/\widehat{m}_0$.

This procedure is based on the observation that the plot of p_i versus i (the quantile plot of the p-values) should exhibit a linear relationship, wherein the slope $S = 1/(m+1)$ passes through (a) the origin and (b) the point $(m+1,1)$ when $m = M_0$. When $M_0 < m$, the p-values corresponding to the false null hypotheses tend to be smaller than the p-values of the true null hypotheses, so they are concentrated on the left side of the plot. The relationship on the right side of the plot remains approximately linear. In this way, the adaptive BH procedure seeks to customise the

Accepted Manuscript

correction procedure based on the estimated M_0, which differs from the traditional BH correction, which provides a universal correction procedure that does not involve the estimation of M_0.

5.3. Results of adaptive Benjamini-Hochberg correction

The techniques of Šidák correction, traditional BH correction and adaptive BH correction are applied to the EDMF back-analyses of the excavation case history. The size of the candidate model set (CMS) is one of the criterion to evaluate the performance of the correction techniques. A smaller CMS is expected to contain fewer Type II errors. Therefore, the size of the CMS reflects the restraints on Type II errors. Figure 11 shows the sizes of the candidate model sets obtained using the three techniques after each round of identification. For the first two rounds of identification, Šidák correction still produces a slightly smaller candidate model set, which is likely due to the smaller number of measurement data involved during these early rounds of identification. However, as more measurement data is included and processed during the 3rd and the 4th round of identification, the adaptive BH correction outperforms both the traditional BH and Šidák correction by yielding the smallest candidate model set.

Figure 12 shows the parallel-axis plot of the identified parameter values using the adaptive BH correction after the 4th round of identification. In contrast to Figure 6, wherein two distinct clusters are observed for the E_{50}^{ref} values of the G(VI) silt layer, the adaptive BH correction produces only one cluster of candidate models. The absence of the second cluster in this case lends support to the earlier postulate that the candidate models in this cluster are erroneous, which are eliminated through the tighter restraint on Type II errors provided by the adaptive BH correction.

The performance of these three correction techniques can be further assessed using the concept of 'statistical power', which is defined as the probability of rejecting a model when it is invalid.

Accepted Manuscript

With reference to Table 6, statistical power is calculated as the ratio $N_{1|1}/(N_{1|1}+N_{0|1})$ (De et al[11]; Farcomeni[13]). Figure 13 shows the normalised statistical power of all three correction techniques across all four rounds of identification. The statistical power of the Šidák correction is adopted as the base for normalisation and therefore, the y-coordinate of the Šidák correction results in Figure 13 is always 1.0. It is noted that the adaptive BH correction produces the highest statistical power, thus demonstrating its improved falsification ability. This is also reflected in its ability to yield the smallest candidate model set shown in Figures 11 and 12.

The effects of these correction techniques on the wall deflection predictions are shown in Figure 14. The two subplots in this figure show, for all measurement depths along wall ID 09 and 04 respectively, the wall deflection range between the computed upper and lower bounds at the final excavation stage, normalized by the corresponding range predicted using Šidák correction. The effect of choosing an alternative correction technique is more profound for wall ID 09. The predicted deflection ranges obtained with the adaptive BH correction are, on average, 40% and 10% smaller than the ranges obtained using the Šidák correction and the traditional BH correction respectively. However, its influence is not so obvious for wall ID 04. This is likely because the deflection magnitudes of wall ID 04 are less sensitive to changes in parameter values.

The primary goal in adopting the BH correction technique is to reduce Type II errors. A closer examination of the candidate model sets obtained using all three correction technics reveals a significant degree of overlap. This means that almost all candidate models obtained with adaptive BH correction are also candidate models obtained using the Šidák correction. Therefore, the overall information on soil parameter values remain consistent across the three correction techniques. The falsification of more model instances by the adaptive BH correction technique, which results in a smaller candidate model set and smaller bounds in the predicted wall deflection, is due to its improved restraint on Type II errors.

Accepted Manuscript

6. CONCLUSIONS

This paper examines in detail the performance of three data-interpretation methodologies (EDMF, Bayesian Model Updating and Residual Minimization) applied to an excavation case history. The enhanced understanding and appreciation on the strengths and limitations of each methodology allow engineers and researchers to choose the most appropriate methodology to obtain reliable and good quality back analysis results. The main conclusions are summarised as follows:

i) The identified parameter values from the three data-interpretation methodologies can differ significantly, based on the current case study. The discrepancies in the identified parameter values and the resulting wall deflection predictions are related to the specific implementation details and assumptions adopted in the data-interpretation methodologies.

ii) In the Bayesian model updating approach implemented in the current study, the choice of correlation matrix significantly affects the values of the parameters that are identified. In the current study, the difference in parameter values between the two correlation schemes can be as high as 90%.

iii) In the Bayesian and residual minimisation analyses performed in the current study, the contributions of some measurement data are not effectively recognised, which adversely affects the accuracy of the predictions. This issue is especially significant when measurement data of inclinometers at two locations are utilised simultaneously.

iv) The residual minimisation method, which indirectly accounts for uncertainties in the form of deterministic weighting terms without considering their distributions, does not perform as well as Bayesian model updating and EDMF in handling

Accepted Manuscript

measurement data near the ground surface, where the uncertainty errors can be twice those present at deeper depths.

v) EDMF, as compared to Bayesian model updating and residual minimisation, yields improved predictions and identified parameter values, especially when measurement data of inclinometers at several locations are simultaneously utilised.

vi) The use of adaptive BH correction in EDMF analysis improves the restraint on Type II errors, as compared to the use of Šidák and the traditional BH correction techniques, which leads to smaller prediction ranges.

Acknowledgments

This research was conducted at the Future Cities Laboratory at the Singapore-ETH Centre (SEC). The SEC was established as a collaboration between ETH Zurich and National Research Foundation (NRF) Singapore (FI 370074011) under the auspices of the NRF's Campus for Research Excellence and Technological Enterprise (CREATE) program.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

REFERENCES

1. Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, *3*, pp.103-107.
2. Ahmadi, M., Vahabzadeh, F., Bonakdarpour, B., Mofarrah, E. and Mehranian, M., 2005. Application of the central composite design and response surface methodology to the advanced treatment of olive oil processing wastewater using Fenton's peroxidation. *Journal of Hazardous Materials*, *123*(1-3), pp.187-195.
3. Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pp.289-300.
4. Benjamini, Y. and Hochberg, Y., 2000. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of educational and Behavioral Statistics*, *25*(1), pp.60-83.
5. Benjamini, Y., Krieger, A.M. and Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), pp.491-507.
6. Benz, T., 2007. *Small-strain stiffness of soils and its numerical consequences* (Vol. 5). Stuttgart: Univ. Stuttgart, Inst. f. Geotechnik.
7. Brinkgreve, R.B.J., Engin, E., Swolfs, W.M., Waterman, D., Chesaru, A., Bonnier, P.G. and Galavi, V., 2012. Plaxis 3D 2012. *Plaxis bv*.
8. Brinkgreve, R.B.J., Kumarswamy, S., Swolfs, W.M., Waterman, D., Chesaru, A. and Bonnier, P.G., 2016. Plaxis 2014. *Delft*, *Netherlands*.
9. Cao, W.J., Koh, C.G. and Smith, I.F.C., 2019. Enhancing static-load-test identification of bridges using dynamic data. *Engineering Structures*, *186*, pp.410-420.
10. Dong, Y.P., Burd, H.J. and Houlsby, G., 2016. Finite-element analysis of a deep excavation case history. *Géotechnique*, *66*(1).

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

11. De, S., Brewick, P.T., Johnson, E.A. and Wojtkiewicz, S.F., 2018. Investigation of model falsification using error and likelihood bounds with application to a structural system. *Journal of Engineering Mechanics*, *144*(9), p.04018078.
12. Finno, R.J. and Calvello, M., 2005. Supported excavations: observational method and inverse modeling. *Journal of Geotechnical and Geoenvironmental Engineering*, *131*(7), pp.826-836.
13. Farcomeni, A., 2008. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical methods in medical research*, *17*(4), pp.347-388.
14. Galavi, V., 2010. Groundwater flow, fully coupled flow deformation and undrained analyses in PLAXIS 2D and 3D. *PLAXIS Internal Research Rep., Plaxis bv, Delft, Netherlands*, *290*.
15. Goulet, J.A., Kripakaran, P. and Smith, I.F.C., 2010. Multimodel structural performance monitoring. *Journal of Structural Engineering*, *136*(10), pp.1309-1318.
16. Goulet, J.A. and Smith, I.F., 2013. Structural identification with systematic errors and unknown uncertainty dependencies. *Computers & structures*, *128*, pp.251-258.
17. Hsein Juang, C., Luo, Z., Atamturktur, S. and Huang, H., 2012. Bayesian updating of soil parameters for braced excavations using field observations. *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(3), pp.395-406.
18. Huang, Z.H., Zhang, L.L., Cheng, S.Y., Zhang, J. and Xia, X.H., 2015. Back-analysis and parameter identification for deep excavation based on Pareto multiobjective optimization. *Journal of Aerospace Engineering*, *28*(6), p.A4014007.
19. Hardy, S., Nicholson, D., Ingram, P., Gaba, A., Chen, Y. and Biscontin, G., 2017, September. New observational method framework and application. In *ICSMGE 2017-*

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

19th International Conference on Soil Mechanics and Geotechnical Engineering (Vol. 2017, pp. 1995-1998).

20. Ieronymaki, E.S., Whittle, A.J. and Sureda, D.S., 2016. Interpretation of free-field ground movements caused by mechanized tunnel construction. *Journal of Geotechnical and Geoenvironmental Engineering*, *143*(4), p.04016114.
21. Ieronymaki, E., Whittle, A.J. and Einstein, H.H., 2018. Comparative study of the effects of three tunneling methods on ground movements in stiff clay. *Tunnelling and Underground Space Technology*, *74*, pp.167-177.
22. Jofré, G.A.C., 2013. Methodology for updating numerical predictions of excavation performance.
23. Jin, Y.F., Yin, Z.Y., Zhou, W.H. and Huang, H.W., 2019. Multi-objective optimization-based updating of predictions during excavation. *Engineering Applications of Artificial Intelligence*, *78*, pp.102-123.
24. Knabe, T., Schweiger, H.F. and Schanz, T., 2012. Calibration of constitutive parameters by inverse analysis for a geotechnical boundary problem. *Canadian Geotechnical Journal*, *49*(2), pp.170-183.
25. Ledesma, A., Gens, A. and Alonso, E.E., 1996. Estimation of parameters in geotechnical backanalysis—I. Maximum likelihood approach. *Computers and Geotechnics*, *18*(1), pp.1-27.
26. Leung, E., Pappin, J. and Koo, R., 2010. Determination of small strain modulus and degradation for in-situ weathered rock and old alluvium deposits.
27. Levasseur, S., Malécot, Y., Boulon, M. and Flavigny, E., 2008. Soil parameter identification using a genetic algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics*, *32*(2), pp.189-213.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

28. Lee, F.H., Hong, S.H., Gu, Q. and Zhao, P., 2010. Application of large three-dimensional finite-element analyses to practical problems. *International Journal of Geomechanics*, *11*(6), pp.529-539.

29. Moon, S.W., Hayashi, K. and Ku, T., 2017. Estimating spatial variations in bedrock depth and weathering degree in decomposed granite from surface waves. *Journal of geotechnical and geoenvironmental engineering*, *143*(7), p.04017020.

30. Moser, G., Paal, S.G. and Smith, I.F.C., 2015. Performance comparison of reduced models for leak detection in water distribution networks. *Advanced Engineering Informatics*, *29*(3), pp.714-726.

31. Ou, C.Y., Chiou, D.C. and Wu, T.S., 1996. Three-dimensional finite element analysis of deep excavations. *Journal of Geotechnical Engineering*, *122*(5), pp.337-345.

32. Peck, R.B., 1969. Advantages and limitations of the observational method in applied soil mechanics. *Geotechnique*, *19*(2), pp.171-187.

33. Popper K. The logic of scientific discovery. Routledge; 2005 Nov 4.

34. Proverbio, M., Costa, A. and Smith, I.F.C., 2018. Adaptive sampling methodology for structural identification using radial-basis functions. *Journal of Computing in Civil Engineering*, *32*(3), p.04018008.

35. Pai, S.G., Nussbaumer, A. and Smith, I.F., 2018. Comparing structural identification methodologies for fatigue life prediction of a highway bridge. *Frontiers in Built Environment*, *3*, p.73.

36. Qi, X.H. and Zhou, W.H., 2017. An efficient probabilistic back-analysis method for braced excavations using wall deflection data at multiple points. *Computers and Geotechnics*, *85*, pp.186-198.

Accepted Manuscript

37. Rahardjo, H., Aung, K.K., Leong, E.C. and Rezaur, R.B., 2004. Characteristics of residual soils in Singapore as formed by weathering. *Engineering Geology*, *73*(1-2), pp.157-169.

38. Rasmussen, C.E., 2004. Gaussian processes in machine learning. In *Advanced lectures on machine learning* (pp. 63-71). Springer, Berlin, Heidelberg.

39. Rahardjo, H., Satyanaga, A., Leong, E.C., Ng, Y.S. and Pang, H.T.C., 2012. Variability of residual soil properties. *Engineering geology*, *141*, pp.124-140.

40. Šidák, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), pp.626-633.

41. Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, *29*(2), pp.143-151.

42. Vernay, D.G., Raphael, B. and Smith, I.F.C., 2014. Augmenting simulations of airflow around buildings using field measurements. *Advanced Engineering Informatics*, *28*(4), pp.412-424.

43. Wang, Z.Z., Goh, S.H., Koh, C.G. and Smith, I.F.C., 2019. An efficient inverse analysis procedure for braced excavations considering three-dimensional effects. *Computers and Geotechnics*, *107*, pp.150-162.

44. Yin, Z.Y., Jin, Y.F., Shen, J.S. and Hicher, P.Y., 2018. Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement. *International Journal for Numerical and Analytical Methods in Geomechanics*, *42*(1), pp.70-94.

45. Zhang, L. L., Zhang, J., Zhang, L. M., & Tang, W. H. (2010). Back analysis of slope failure with Markov chain Monte Carlo simulation. *Computers and Geotechnics*, *37*(7-8), 905-912.

Accepted Manuscript

46. Zdravkovic, L., Potts, D.M. and St John, H.D., 2011. Modelling of a 3D excavation in finite element analysis. In *Stiff Sedimentary Clays: Genesis and Engineering Behaviour: Géotechnique Symposium in Print 2007* (pp. 319-335). Thomas Telford Ltd.

47. Zhang, W.G., Zhang, R.H., Han, L. and Goh, A.T.C., 2019. Engineering properties of the Bukit Timah Granitic residual soil in Singapore. *Underground Space*, *4*(2), pp.98-108.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

Table 1 Properties of structural elements involved in the excavation case history.

	Diaphragm Walls	Toe Pins	Soldier Pile Walls	Concrete Waler Beams Type 1/2/3	Steel Waler Beams Type 1/2	Struts
Thickness (m)	0.8	-	-	-	-	-
EA(kN)	2.0E7	18E6	-	1.7E7/3.2E7/2.48E7	4.0E6/1.3E7	8.0E6
EI(kNm^2)	1.1E6	11E3	**(3000-10000)**	7.0E5/2.1E6/1.3E6	2.6E5/8.8E5	-
$L_{spacing}$ (m)	-	-	-	-	-	10

Table 2 Properties of geological materials involved in the excavation case history.

	Fill	Gravel	Sandy Silt Residual Soil	Coarse Sand	Rock
E (MPa)	**(3-20)**	40	-	-	2.5E3
E_{50}^{ref}(MPa)	-	-	**(5-50)**	**(5-50)**	-
E_{oed}^{ref}(MPa)	-	-	***1.0 * E_{50}^{ref}***	***1.0 * E_{50}^{ref}***	-
E_{ur}^{ref}(MPa)	-	-	***3.0 * E_{50}^{ref}***	***3.0 * E_{50}^{ref}***	-
m	-	-	0.6	0.6	-
c' (kPa)	0	0	10	0	-
φ'(o)	25	30	28	35	-
ψ(o)	0	0	0	0	-
$\Upsilon_{0.7}$	-	-	0.0001	0.0001	-
G_0^{ref}(MPa)	-	-	***2 * E_{ur}^{ref}***	***2 * E_{ur}^{ref}***	-
p^{ref}(MPa)	-	-	100	100	-
σci (MPa)	-	-	-	-	80
mi	-	-	-	-	32.7
GSI	-	-	-	-	65
D	-	-	-	-	0.7
R_{inter}	0.7	0.5	0.7	0.7	0.75
k (m/s)	2E-6	2E-5	5E-7	2E-6	2E-6

Accepted Manuscript

Table 3 Simplified excavation activities and remarks.

Stage	Simplified Excavation Activities	Duration (days)	Calculation Type
0A	Initial Condition	-	Gravity Loading
0B	Wall Installation	-	Plastic
1	Excavate below Strut layer 1	20	Fully coupled flow-deformation
2	Install Strut layer 1	45	Fully coupled flow-deformation
3	Excavate below Strut layer 2	20	Fully coupled flow-deformation
4	Install Strut layer 2 and Excavate to formation level	30	Fully coupled flow-deformation

Table 4 Examples of uncertainties used in the study.

Uncertainty sources	Magnitudes	Remarks
Inclinometer uncertainties	e.g. ±3.5mm	Finno and Calvelo[12]
2D-model simplification	e.g. 0.9mm – 2.3mm	Wang et al[43]
Response surface	e.g. ±2.5mm	-
FEM	±5%	Goulet et al[15]
Others	±5%	Goulet et al[15]

Table 5 Correlation coefficient at selected depth.

Reduced Level (m)	Partial matrix of correlation coefficient computed from Equation 6 (Qi and Zhou[36] and Ledesma et al[25])						
122	1.00	0.98	0.90	0.78	0.65	0.46	0.23
120	0.98	1.00	0.97	0.90	0.80	0.63	0.40
118	0.90	0.97	1.00	0.98	0.91	0.78	0.56
116	0.78	0.90	0.98	1.00	0.98	0.89	0.70
114	0.65	0.80	0.91	0.98	1.00	0.97	0.83
112	0.46	0.63	0.78	0.89	0.97	1.00	0.94
110	0.23	0.40	0.56	0.70	0.83	0.94	1.00

Table 6 Possible outcomes of a multiple hypothesis testing exercise.

	H_0 is accepted	H_0 is rejected	Total
H_0 is true	$N_{0\|0}$	$N_{1\|0}$ (Type I)	M_0
H_0 is false	$N_{0\|1}$ (Type II)	$N_{1\|1}$	M_1
Total	m-R	R	m

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

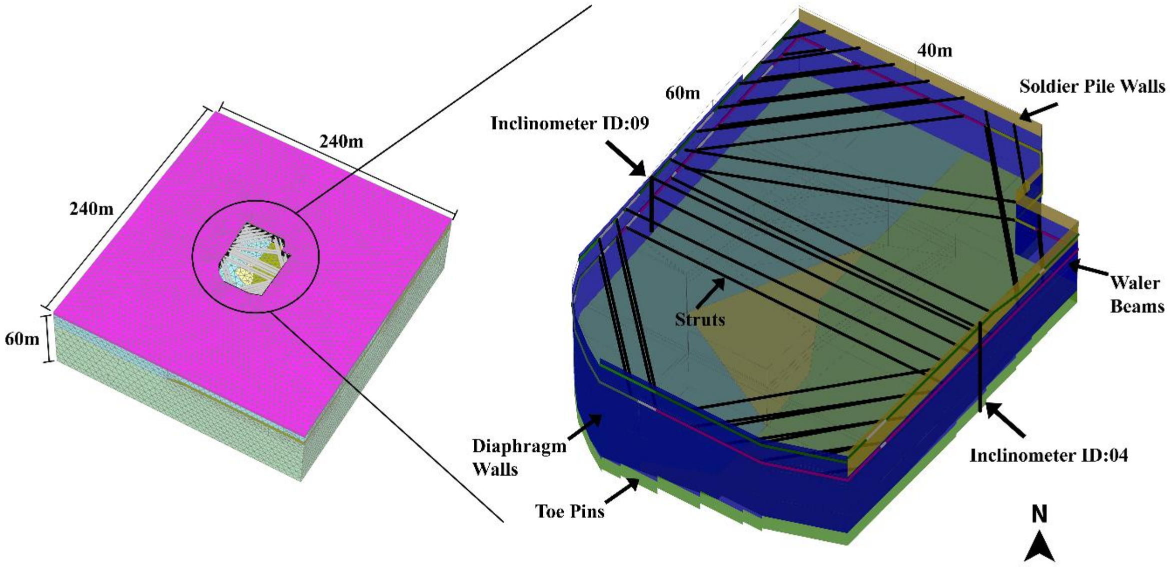

Figure 1 Three-dimensional FEM model of the excavation case history. With zoom-in view of the excavation pit and support system.

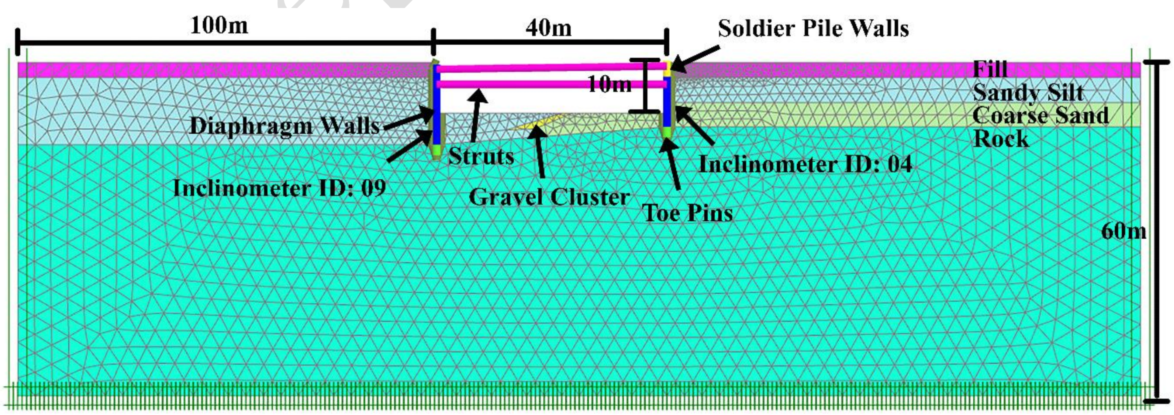

Figure 2 Two-dimensional FEM model of the west-to-east section of the excavation case history.

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

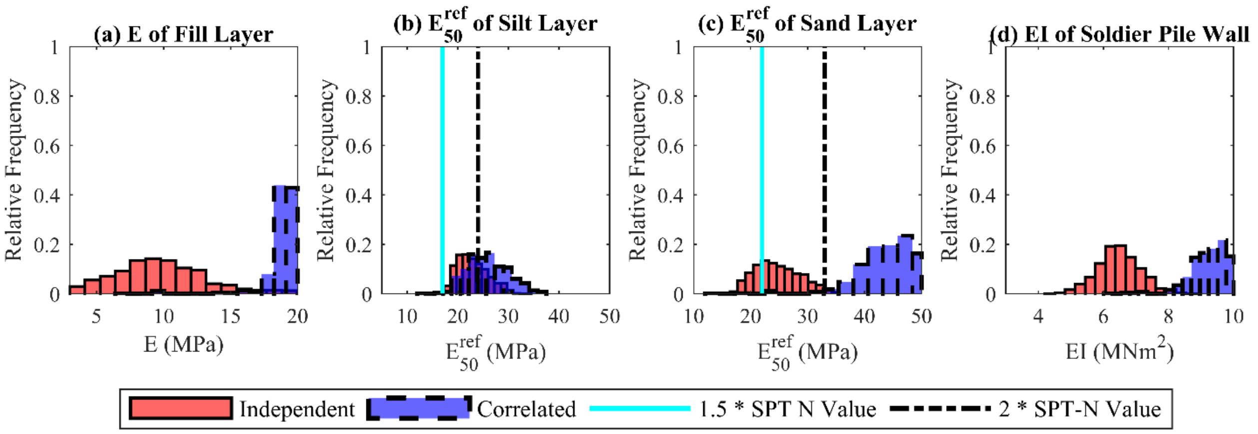

Figure 3 Posterior distributions of the four parameters after the 4th round of identification, from Bayesian model updating analysis.

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

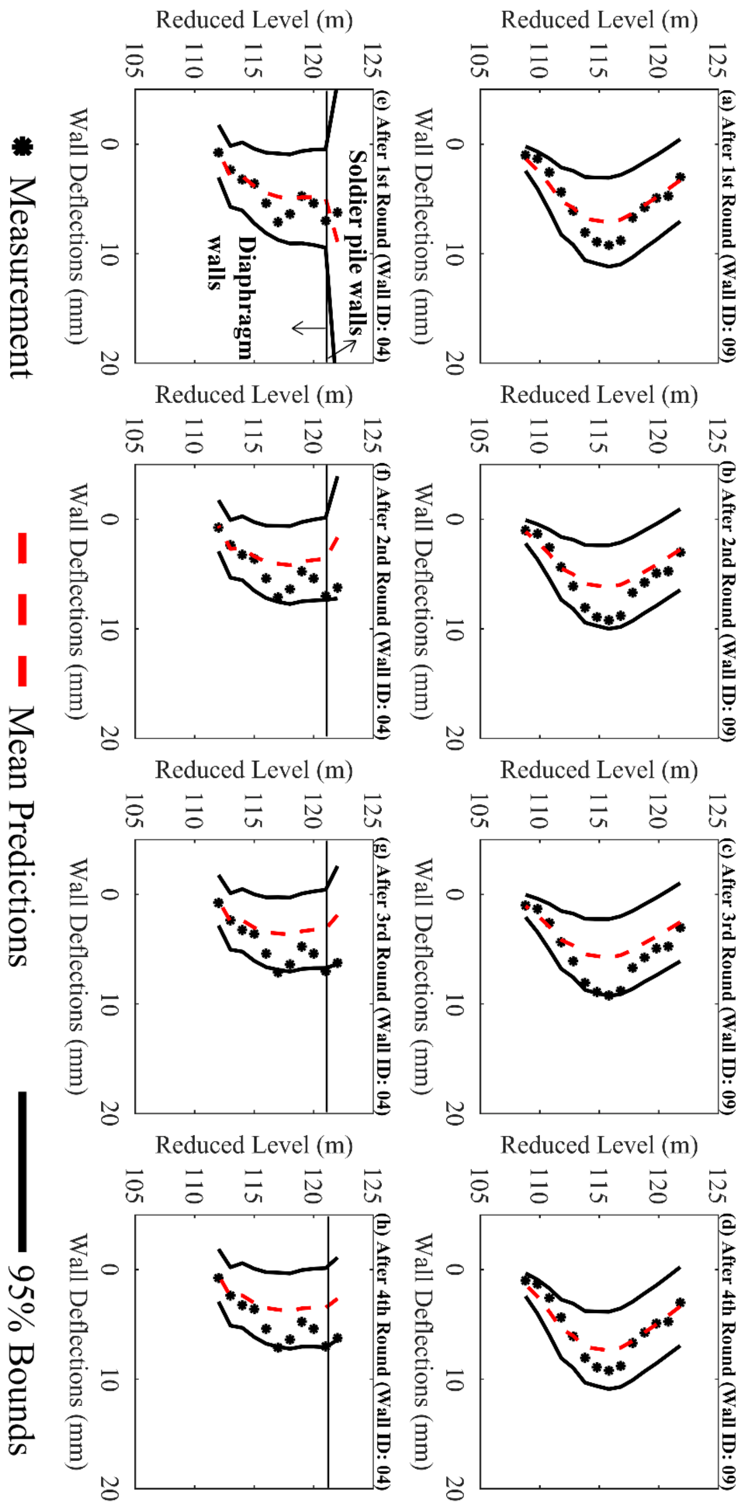

Figure 4 Measured vs predicted (mean and 95% bounds) wall deflections of the final excavation stage after each round of identification, using Bayesian model updating with non-zero correlation scheme. (Top row: ID 09 and bottom row: ID 04)

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

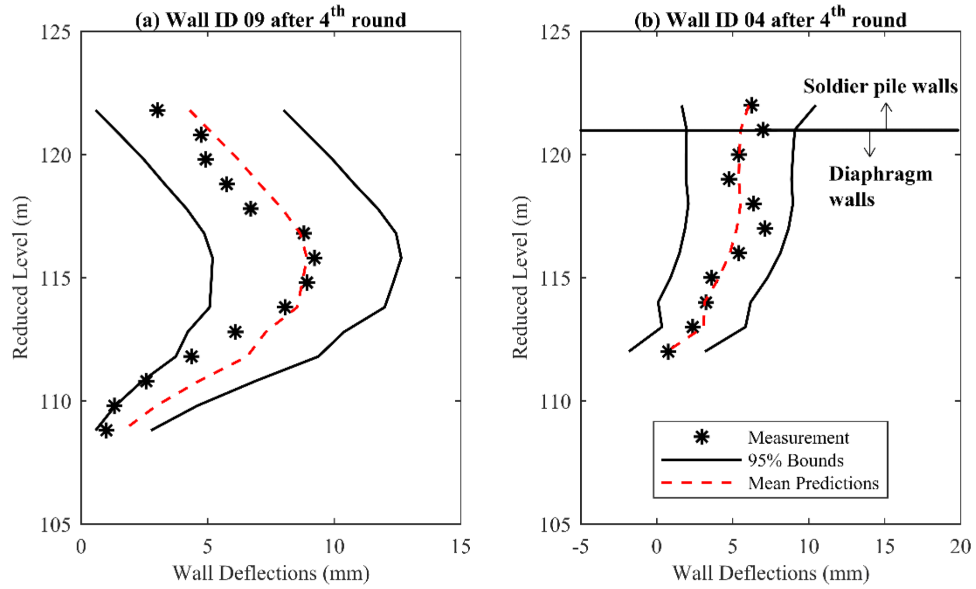

Figure 5 Measured vs predicted (mean and 95% bounds) wall deflections of the final excavation stage after the 4th round of identification, using Bayesian model updating with non-zero correlation scheme (with measurements of ID 04 only).

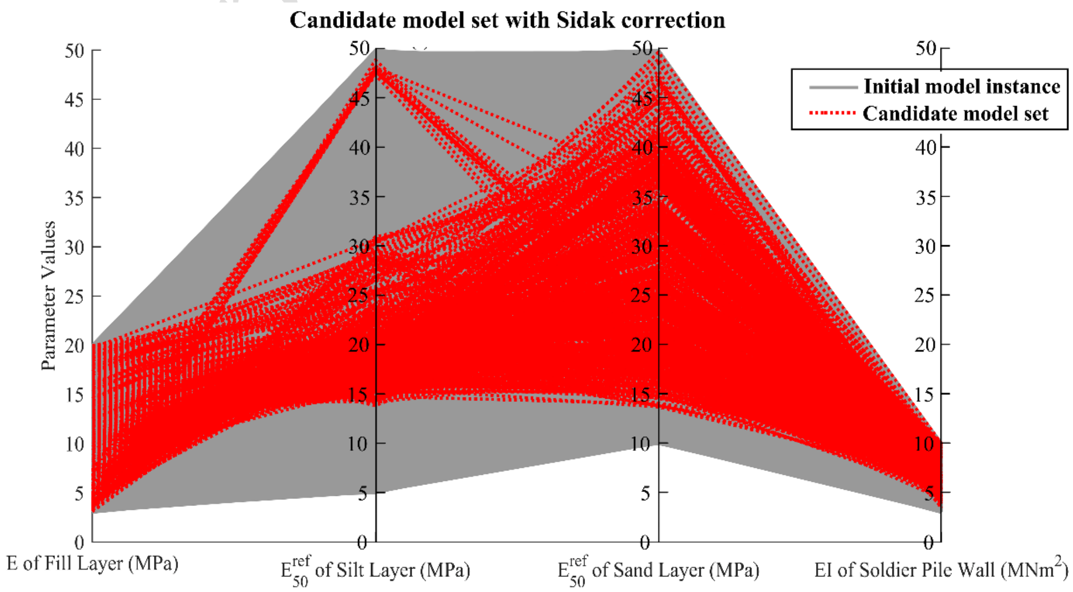

Figure 6 Parallel-axis plot of the identified parameter values from EDMF analysis with Šidák correction after the 4th round of identification. Each red dashed line connects the identified parameter corresponding to a candidate model.

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

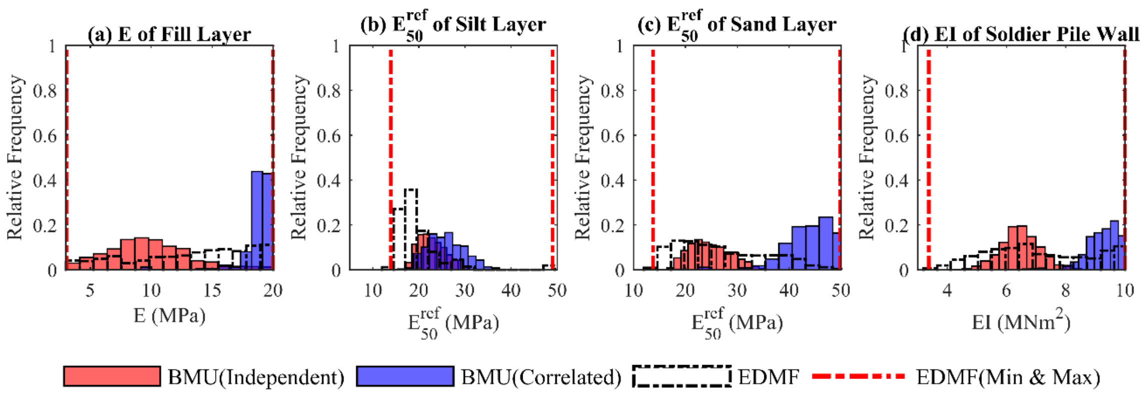

Figure 7 Comparison of the distribution of parameter values identified using BMU and EDMF after the 4th round of identification. (BMU: Bayesian model updating. EDMF: error-domain model falsification).

Accepted Manuscript

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

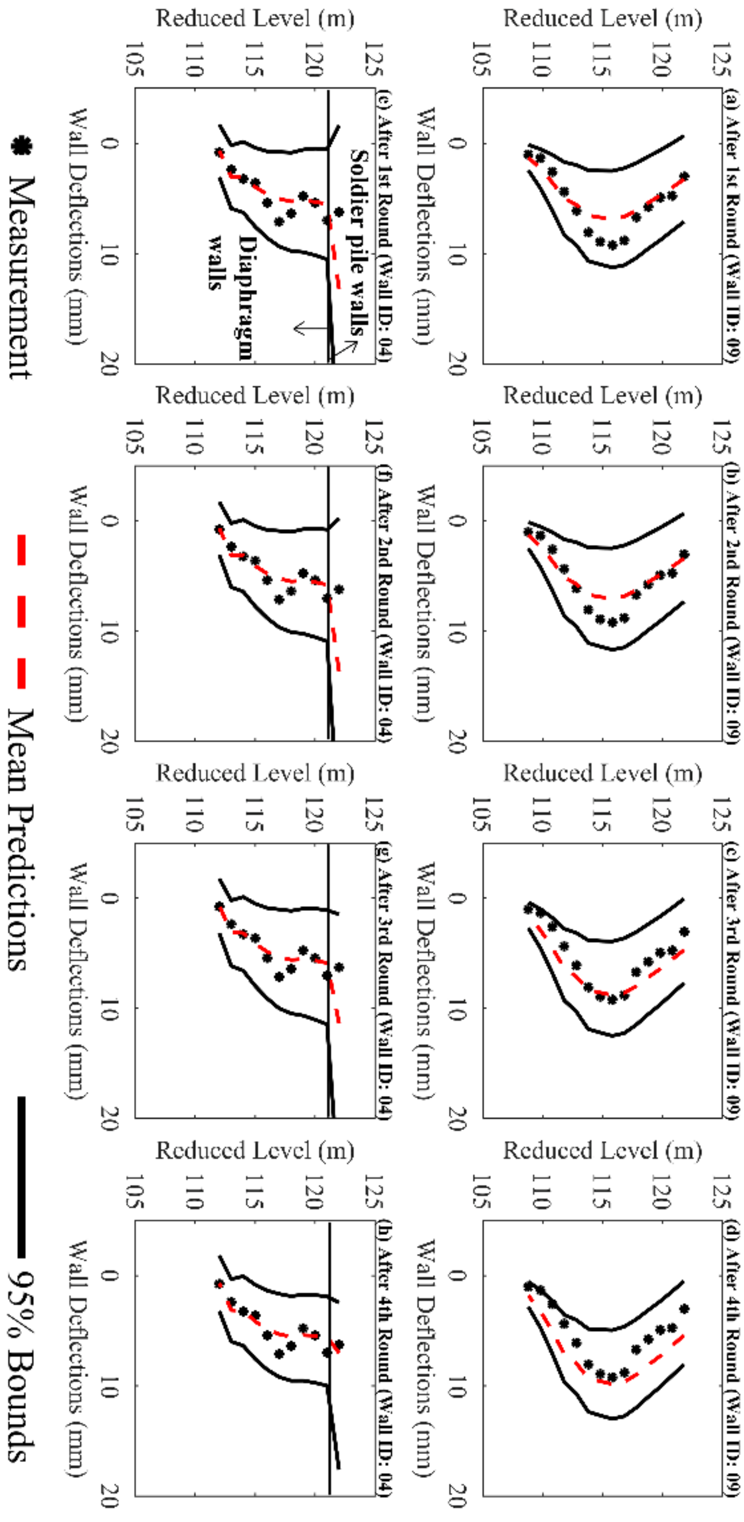

Figure 8 Measured vs predicted (mean and 95% bounds) wall deflections of the final excavation stage after each round of identification, using EDMF. (Top row: ID 09 and bottom row: ID 04)

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

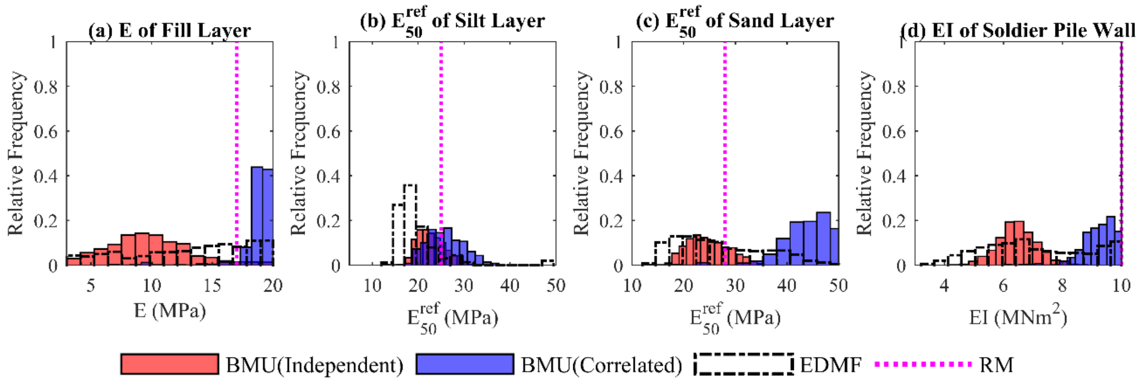

Figure 8 Comparison of distribution of parameter values identified using BMU, EDMF and RM after the 4th round of identification. (BMU: Bayesian model updating. EDMF: error-domain model falsification. RM: residual minimisation). The RM result is a single value shown by the purple dashed line.

Accepted Manuscript

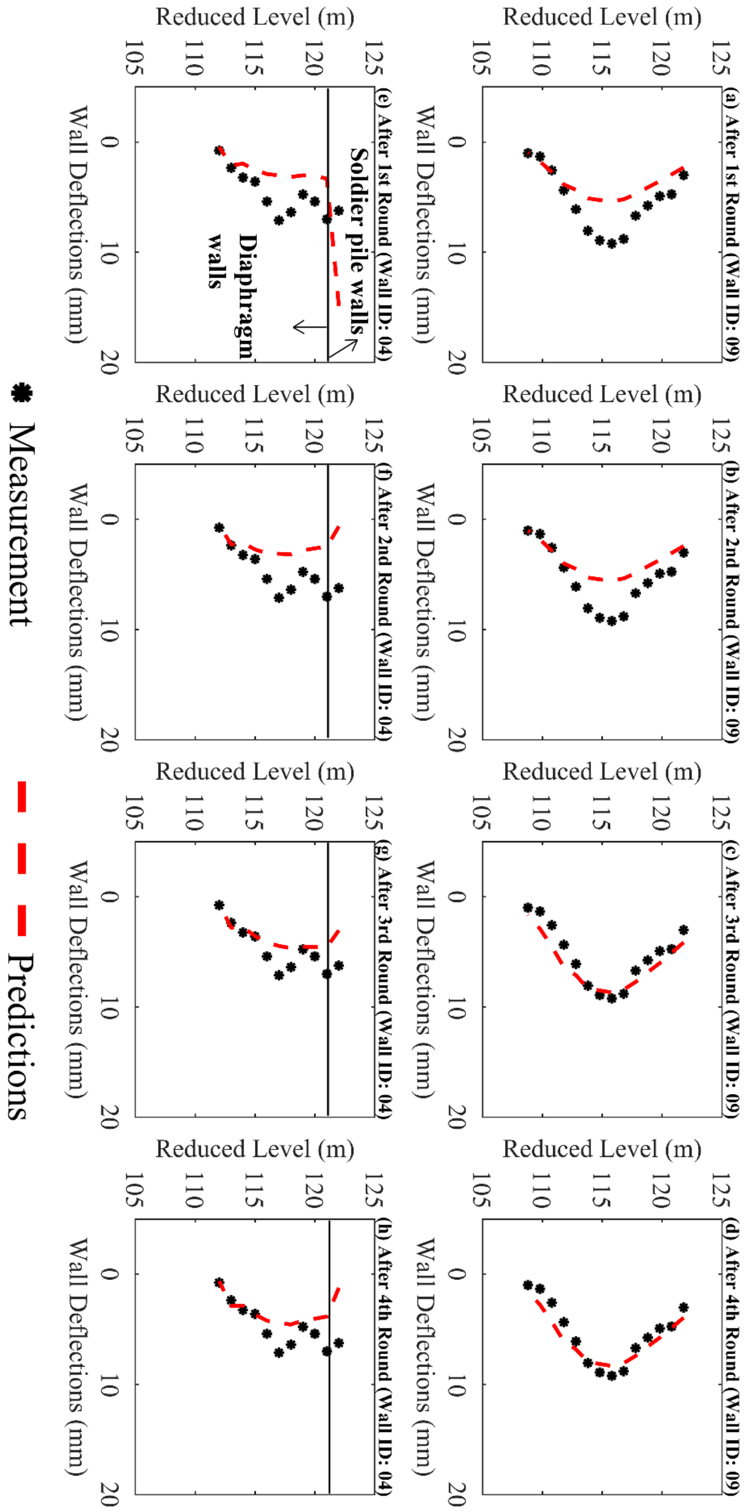

Figure 10 Measured vs predicted wall deflections of the final excavation stage after each round of identification, using residual minimisation. (Top row: ID 09 and bottom row: ID 04)

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

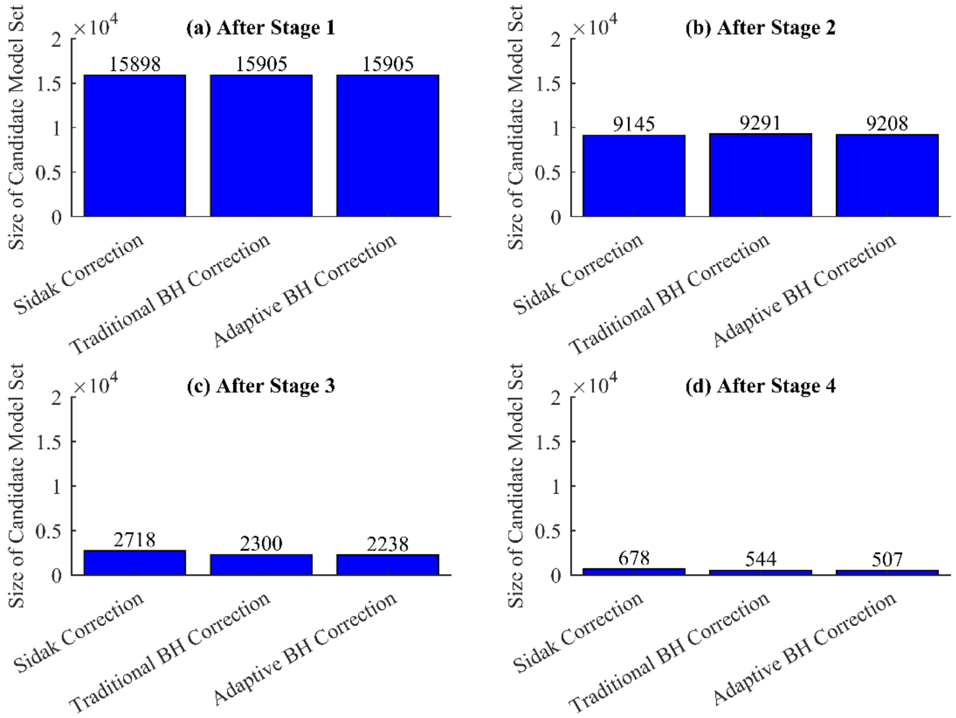

Figure 9 Comparisons of the size of the candidate model set obtained using three correction techniques, after each round of identification.

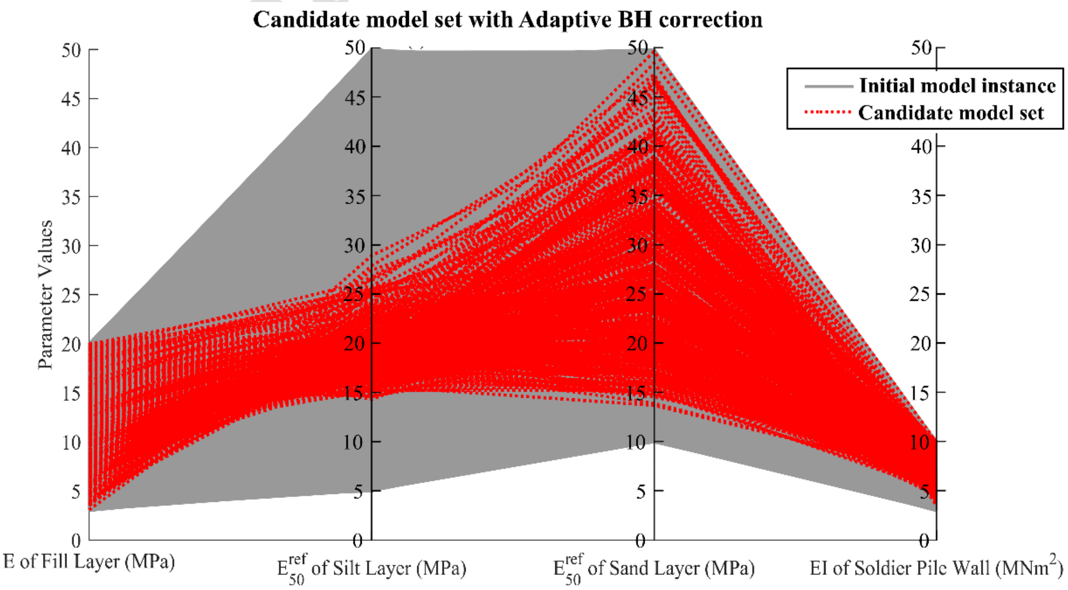

Figure 10 Parallel-axis plot of the identified parameter values from EDMF analysis with adaptive BH correction after the 4th round of identification.

Wang, Z. Z., Goh, S. H., Koh, C. G., & Smith, I. F. (2020). Comparative study of the effects of three data-interpretation methodologies on the performance of geotechnical back analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, *44*(15), 2093-2113. https://doi.org/10.1002/nag.3120

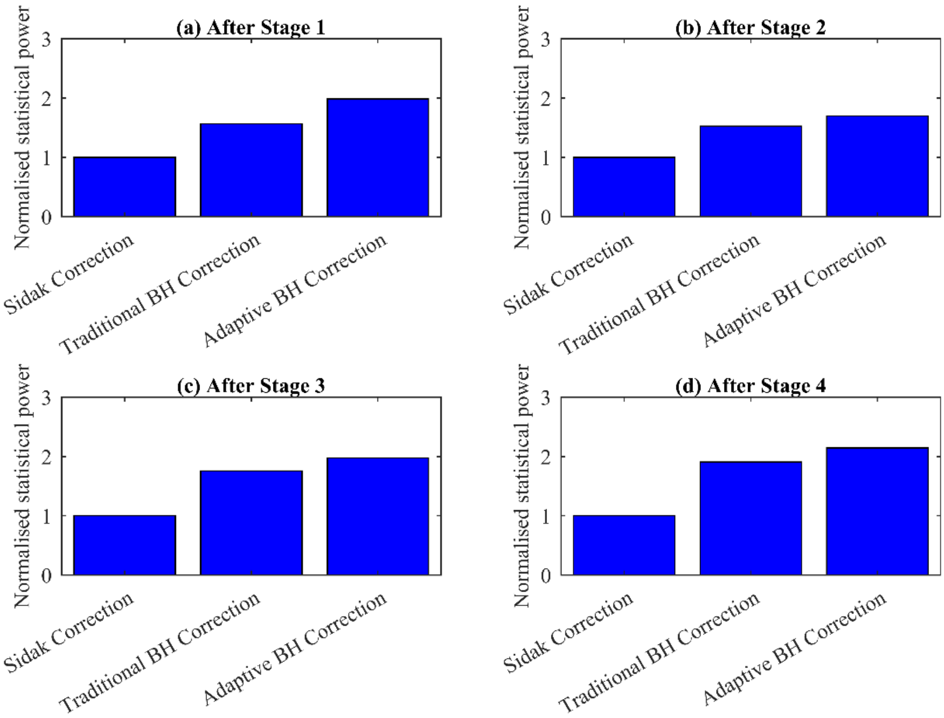

Figure 11 Comparison of the statistical power associated with the three correction techniques, after each round of identification.

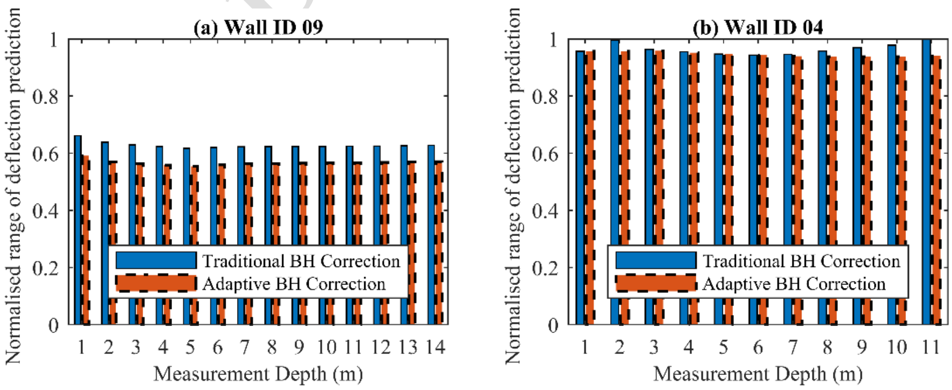

Figure 12 Comparison of the normalized wall deflection range between 95% bounds after the 4th round of identification, obtained using the three correction techniques. The results are normalised with respect to those obtained using Sidak correction.