MULTI-SOURCE DATA INTEGRATION FOR SOIL PARAMETER UNCERTAINTY QUANTIFICATION

A PREPRINT

Ningxin Yang*
Department of Civil
and Environmental Engineering
Imperial College London
London, UK

Truong Le
Department of Civil
and Environmental Engineering
Imperial College London
London, UK

Lidija Zdravkovic
Department of Civil
and Environmental Engineering
Imperial College London
London, UK

David Potts
Department of Civil
and Environmental Engineering
Imperial College London
London, UK

September 2, 2026

ABSTRACT

Calibration of soil parameters is essential for reliable geotechnical analysis, yet remains challenging due to limited field measurements, model discrepancy, correlated outputs, and the need to incorporate multiple-source data types. This paper proposes a sequential Bayesian framework that enables staged data assimilation and progressive refinement of a wide range of model parameters, thereby improving the reliability of predictions of engineering behaviour throughout the entire construction and operation life cycle. A reduced-order modelling approach, combining principal component analysis and polynomial chaos expansion, is employed to construct efficient surrogates that preserve output correlations and support global sensitivity analysis. A custom likelihood formulation is developed to integrate multiple observation types while explicitly accounting for model discrepancy and uncertainty. To focus calibration efforts on critical regions, a spatial weighting strategy is used, reflecting zones of engineering significance. The proposed methodology is demonstrated through a full-scale laterally loaded pile test. Results indicate that the surrogate model yields accurate and

*corresponding author: n.yang23@imperial.ac.uk

computationally efficient predictions, and that the Bayesian updating process systematically reduces parameter uncertainty while maintaining consistency with field observations.

Keywords Sequential Bayesian inference; Uncertainty quantification; Reduced-order modeling; Model discrepancy; Multi-source data;

1 Introduction

Recent advances in field monitoring and numerical modelling have enabled inverse analyses that infer material parameters from site observations. Techniques are being developed to overcome challenges associated with large volumes of monitoring data and the computational cost of numerical modelling. Of the proposed methods, Bayesian methods are well suited to situations where data are limited and engineering judgement is integral, offering a probabilistic framework for integrating prior knowledge and quantifying uncertainty. (Kelly and Huang 2015; Wang et al. 2016; Baecher 2017; Rahimi et al. 2019; Lo and Leung 2019; Jin et al. 2021; Yuen et al. 2021; Zhou et al. 2021; Buckley et al. 2023). Sequential Bayesian updating is particularly attractive in staged construction, where parameters can be progressively refined as new observations become available (Qi and Zhou 2017; Jin et al. 2021; Tang et al. 2022).

Unfortunately, inverse analysis often demands extensive forward simulations, making uncertainty quantification (UQ) computationally prohibitive. Surrogate models reduce this computational cost by approximating high-fidelity input-output relationships; however, many do not adequately capture the spatial correlations that are critical in geotechnical systems (Qi and Zhou 2017; Lo and Leung 2019; Westermann and Evins 2019; Zheng et al. 2023a; Zheng et al. 2023b). Reduced order modelling addresses this limitation by projecting system responses onto lower dimensional basis that preserve the key physics (Blatman and Sudret 2011; Peherstorfer and Willcox 2016; Brunton et al. 2020; McQuarrie et al. 2021; Brunton and Kutz 2022; Geelen et al. 2023; Kramer et al. 2024; Conti et al. 2024). Geotechnical analyses can produce outputs with many degrees of freedom, such as nodal displacements and pore-water pressures at large numbers of mesh nodes and, where relevant, across multiple load increments or excavation stages. Despite this high output dimensionality, the responses commonly exhibit a correlated spatial structure that can be described using a small number of dominant spatial patterns (Brunton and Kutz 2022). These patterns are often termed modes; each mode represents a characteristic spatial distribution of displacement obtained from the numerical outputs. Dimensionality reduction identifies the patterns associated with the largest variation, ranks them by their contribution to the total variation, and thereby selects the dominant modes. The modes may be interpreted as the principal structures that govern the magnitude and spatial distribution of the geotechnical outputs. A surrogate model defined in this reduced space can then be trained and evaluated to reproduce the essential behaviour of the original model at a reduced computational training cost. This approach enables efficient handling of vector-valued, correlated responses and global sensitivity analysis (Sudret 2008; Garcia-Cabrejo and Valocchi 2014; Nagel et al. 2020; Wagner et al. 2020; Yang et al. 2026). First, this study uses *principal component analysis* (PCA) and *polynomial chaos expansion*

(PCE) to build the surrogate that captures interdependencies among high-dimensional outputs and enables efficient and analytical sensitivity analysis.

While computer models and monitoring systems encompass multiple quantities of interest, most Bayesian calibration works rely on single-source data, e.g., using displacements while neglecting internal forces (Wang et al. 2016; Qi and Zhou 2017; Jin et al. 2021). This fragmentation limits uncertainty quantification and parameter identifiability. Notable efforts have been made to merge multiple-source monitoring data for predictive purposes (Kelly and Huang 2015; Rahimi et al. 2019; Tian et al. 2024), unfortunately many lack probabilistic frameworks or sequential updates. A second contribution of this study is therefore the development of a joint-likelihood-based function incorporating multiple relevant observations and model quantities of interest. Since calibration outcomes are sensitive to how uncertainties from both data and model predictions are represented, particular attention is given to the formulation of this likelihood function.

Accurate likelihood construction in inverse analysis requires explicit consideration of multiple uncertainty sources, including *observation noise*, *surrogate error*, and often-ambiguous model discrepancies, which together contribute to the differences between observations and predictions (Kennedy and O'Hagan 2001). In the geotechnical context, these discrepancies can arise from finite element (FE) model simplifications (*model error*) or from inherent biases in field data acquisition (*systematic error*), such as sensor misalignment or installation issues. While conceptually distinct, these uncertainties are often difficult to untangle, complicating model calibration. Surrogate approximations further introduce additional structural uncertainty to the calibration process. The Bayesian framework of Kennedy and O'Hagan (2001) offers a formal treatment of model-data discrepancy but assumes that observations are inherently more reliable than model predictions (Brynjarsdóttir and O'Hagan 2014; Ling et al. 2014; Rahimi et al. 2019). This assumption should be treated with caution in geotechnical applications, as observations may also be affected by measurement errors and environmental effects that are not explicitly represented in the model. Enforcing agreement between model outputs and observations without resolving the source of discrepancy may lead to overfitting and physically implausible results. To address these, as the third contribution of this study, the Kennedy and O'Hagan (2001) framework is adapted using a *weighting* strategy, which assigns higher weights to regions of geotechnical significance. The weighting is defined prior to calibration based on engineering judgement, with the regions of interest specified at the outset. In addition to emphasising zones of engineering relevance, the spatial weighting also reduces the influence of parts of the domain where model discrepancy is observed. This enables calibration to focus on critical areas while avoiding the effect of the discrepancy between data and model. The custom likelihood function also incorporates *observation noise* and *surrogate error* as explicit uncertainty terms, considering their effects where appropriate. This allows all key uncertainty sources to be explicitly incorporated in the calibration, improving interpretability and maintaining consistency with the physical real-world system.

The proposed framework is demonstrated on a full-scale laterally loaded pile test. A *principal component analysis-polynomial chaos expansion* surrogate is adopted to represent the correlated lateral deflections across loading stages and to enable efficient model evaluations. A custom likelihood function is formulated to assimilate pile displacements

and applied forces jointly, with explicit treatment of observation noise, surrogate error and model discrepancy. Calibration is spatially weighted toward the maximum pile deflection regions to emphasise physical relevance and mitigate the effects of various discrepancies. Sequential Bayesian updating is then applied to fuse the multi-source data across stages, yielding parameter posteriors that reflect the cumulative information content. Results show close agreement between predicted and measured responses, and are physically interpretable when compared with the real site data.

2 Probabilistic model

2.1 Uncertainty

Uncertainty in geotechnical modelling arises from both inherent system variability and incomplete knowledge, typically categorised as aleatory and epistemic uncertainty, respectively (Christian 2004; Baecher 2023). While aleatory uncertainty (e.g., heterogeneity of noise) is irreducible yet characterisable through more data collection, epistemic uncertainty-stemming from model form assumptions, idealisations, or limited site information, is often more influential in practice. The distinction between the two is not always clear, and engineering judgment is required in managing the associated trade-offs. A taxonomy of typical uncertainty is presented below, highlighting their origins and implications.

- *Forward parameters* x: Reflect epistemic uncertainty due to limited knowledge of key geotechnical inputs, e.g., distribution of soil properties and boundary conditions, many of which are indirectly inferred or unmeasurable.
- *Observation noise* ε_r: Represents aleatory uncertainty from sensor limitations and random experimental effects. In high-dimensional settings, this noise may exhibit a spatial/temporal correlation and heteroscedasticity (where the noise structure may vary across space) (Brynjarsdóttir and O'Hagan 2014).
- *Model discrepancy* ξ: Represents the mismatch between model predictions $\mathcal{M}(x)$ and observations $\mathcal{Y}$, arising from unmodelled physics, numerical inaccuracies, or sensor bias, including *model error* (δ), *systematic error* (s), or a combination of both.
- *Surrogate error* T: Introduced by using a surrogate mathematical approximation $\tilde{\mathcal{M}}$ instead of the underlying numerical physical model $\mathcal{M}$, typically due to model reduction, limited training coverage, truncation, or compression. It is quantified as:

$$T = \mathcal{M}(x) - \tilde{\mathcal{M}}(x), \tag{1}$$

Distinguishing between different sources of uncertainty is essential for effective uncertainty quantification (Oberkampf et al. 2002). When conventional conservative approaches are insufficient, probabilistic approaches such as Bayesian inference or the observational method are often adopted. In this work, uncertainty is quantified within such a probabilistic framework. In particular, the forward parameters x and elements in the observation noise term ε_r are treated as random variables to represent epistemic and aleatory uncertainty, respectively, while model discrepancy ξ and surrogate error T are characterised through error terms derived from the available simulations and observations.

2.2 Model formulation

To relate the model predictions $\mathcal{M}(x)$ to the observations $y \in \mathbb{R}^N$, the observation noise $\varepsilon_r \in \mathbb{R}^N$ is commonly introduced as an additive term accounting for the difference:

$$y = \mathcal{M}(x) + \varepsilon_r, \tag{2}$$

where $x \in \mathbb{R}^M$ represents the uncertain input soil parameters (M is the size of input parameter space). When m independent observations are available, the relationship can be expressed as:

$$\mathcal{Y} = \{y^{(i)}\}_{i=1}^m = \{[y_1^{(i)}, y_2^{(i)}, \cdots, y_N^{(i)}]\}_{i=1}^m = \{\mathcal{M}(x) + \varepsilon_r^{(i)}\}_{i=1}^m, \quad \text{where} \quad \mathcal{Y} \in \mathbb{R}^{m \times N}, \tag{3}$$

where N is the length of each observation vector. In this formulation, the forward FE model $\mathcal{M}(x)$ is deterministic, producing a unique response for a given set of input parameters x. The probabilistic framework is introduced through the soil parameters and observation noise, which are treated as uncertain. As a result, the analysis considers a range of plausible parameter values, leading to a corresponding range of possible model responses for comparison with field measurements. This contrasts with a conventional deterministic analysis, in which a single set of soil parameters is adopted.

Equation (2) assumes a numerical FE model $\mathcal{M}(x)$ accurately reflects the real-world problem, which may not be the case in practice. Beyond ε_r, additional uncertainties (see Section 2.1) should also be considered to avoid a biased inference or poor predictions. In line with the Kennedy and O'Hagan (2001) framework, a comprehensive probabilistic model should incorporate a reliable forward model $\mathcal{M}$ (or a surrogate $\tilde{\mathcal{M}}$) and explicitly account for the principal sources of uncertainty, thereby ensuring a complete and consistent formulation:

$$\begin{aligned} \mathcal{Y} &= \mathcal{M}(x) + \xi + \varepsilon_r \\ &= \tilde{\mathcal{M}}(x) + T + \xi + \varepsilon_r, \quad \text{where} \quad \xi \in \{\emptyset, \delta, s, \delta + s\}, \end{aligned} \tag{4}$$

where ξ represents the model discrepancy and may take one of four forms: $\emptyset$, δ, s, or $\delta + s$. Here, empty set $\emptyset$ denotes the case where the discrepancy is considered negligible, i.e. the FE model response is regarded as sufficiently consistent with the measurements for the purpose of calibration. The model error δ and the systematic error s may act individually or in combination. The surrogate error T can be quantified from Equation (1). ε_r is associated with sensor accuracy, which may be specified by technical specifications, inferred from long-term monitoring records, or treated as unknowns during calibration.

A key challenge lies in distinguishing whether model discrepancy arises from *model error* δ or *systematic error* s. This interpretation directly influences the choice of calibration strategy. As illustrated in Figure 1, two Kennedy and O'Hagan (2001)-type strategies are commonly encountered: adjusting the model to match the data (data-centric), or modifying the data to align with the model (model centric). The choice of either method of correction is often heuristic and reliant on subjective judgment. In geotechnical engineering, such approaches are further limited by weak

contextual relevance and poor physical interpretability. Given that not all regions contribute to engineers' decision-makings, calibration efforts should prioritize zones of practical significance, such as maximum wall deflections during excavation or settlements beneath structures. Guided by this principle, this study adopts a targeted calibration strategy using *spatial weighting*, where discrepancy parts are down-weighted and critical regions are emphasised. As shown in Figure 1 (balanced method), this approach aims to enhance physical fidelity by focusing on regions of engineering relevance, avoiding arbitrary corrections and enabling more meaningful inference.

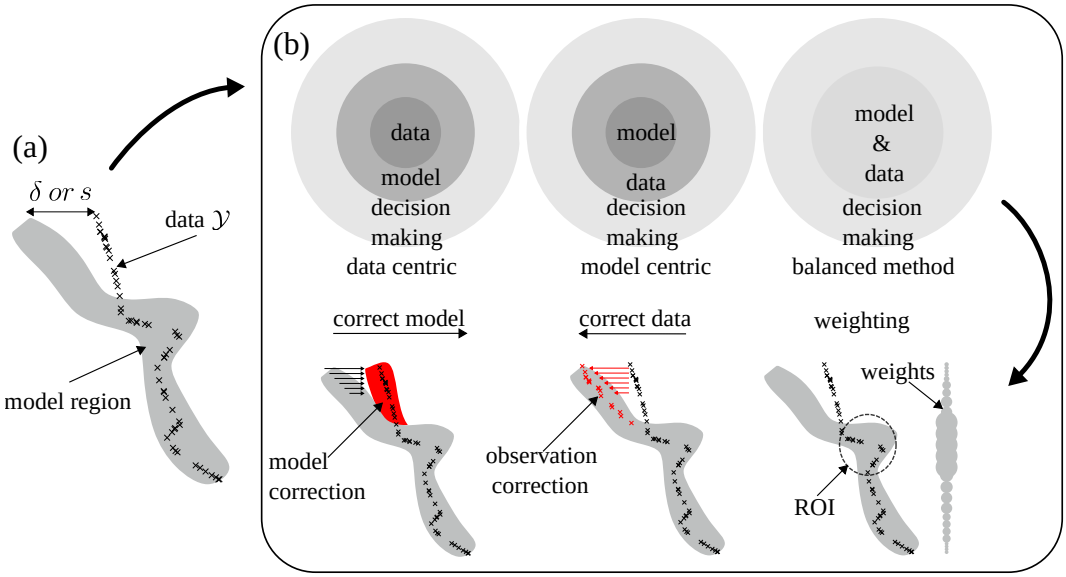


Figure 1: (a) model-data discrepancy; (b) data-centric, model-centric and weighting approaches. Note: ROI = region of interest.

In inverse problems, the uncertain parameter vector is denoted as (x, ε_r), with sensor noise ε_r treated as a variable whose level and variation with depth are inferred during calibration. The *weighting* method simplifies the problem by bypassing the need for explicit distinction between discrepancies. The inverse problem is then formulated accordingly through *Bayes' theorem* as:

$$\pi(x, \varepsilon_r|\mathcal{Y}) = \frac{\mathcal{L}(x, \varepsilon_r|\mathcal{Y}) \cdot \pi(x) \cdot \pi(\varepsilon_r)}{\pi(\mathcal{Y})} \propto \mathcal{L}(x, \varepsilon_r|\mathcal{Y}) \cdot \pi(x) \cdot \pi(\varepsilon_r), \tag{5}$$

where $\pi(x, \varepsilon_r|\mathcal{Y})$ represents the updated estimates of the soil parameters and measurement uncertainty after incorporating the monitoring data $\mathcal{Y}$. The priors $\pi(x)$ and $\pi(\varepsilon_r)$ denote the initial assumptions on the soil parameters and measurement uncertainty, while the likelihood $\mathcal{L}(x, \varepsilon_r|\mathcal{Y})$ measures the agreement between model predictions and field measurements.

Equation (5) typically requires a sampling algorithm, such as Markov Chain Monte Carlo, to approximate the posterior distribution, due to the intractability of computing the evidence $\pi(\mathcal{Y})$ arising from model complexity or high computational cost. To accelerate the calibration process, a polynomial chaos expansion surrogate model (i.e. a mathematical approximation) is used in lieu of the adopted physical model. The Markov chains were constructed using prior knowledge and sampling from an initial or inferred soil parameter distribution. Each chain follows a random

walk based on the Metropolis-Hastings algorithm (Hastings 1970). In each iteration, a candidate sample is generated, and the surrogate model is used to predict the corresponding response. The predicted result is then compared to field observations. This iterative process refines the exploration of the parameter space.

However, as noted above, surrogate modelling introduces errors, T. These errors can be accounted for during inversion by incorporating the error within the *observation noise* covariance Σ_{ε_r} (Hansen et al. 2014; Meles et al. 2022). A total of K realizations of x are generated, each producing N vector outputs used for surrogate training. After constructing the surrogate, the *surrogate error* T is evaluated using a validation dataset of K' independent FE runs, which are not used in the training process, resulting in a $K' \times N$ matrix, $T \in \mathbb{R}^{K' \times N}$. The surrogate error is included in the likelihood as an additional uncertainty contribution. Since the surrogate replaces the original finite-element model, approximation errors in its predictions should not be absorbed into the inferred soil parameters. The observation covariance is therefore augmented by a surrogate-error covariance term by $\alpha_T \in \mathbb{R}^{N \times N}$. This prevents overconfident updating when surrogate errors are non-negligible, while leaving the likelihood essentially unchanged when the surrogate is accurate. The combined covariance is expressed as:

$$\Sigma_T = \frac{1}{N} T^{\mathsf{T}} T, \tag{6}$$

$$\Sigma_{\varepsilon_r T} = \alpha_T \circ \Sigma_{\varepsilon_r}, \tag{7}$$

$$\alpha_T = \mathbf{I} + \sqrt{\mathrm{diag}(\Sigma_T)}/|\bar{\mathcal{Y}}|, \tag{8}$$

where $\circ$ denotes the element-wise (Hadamard) product. The vector $\bar{\mathcal{Y}} \in \mathbb{R}^N$ represents the mean across the m datasets, i.e., $\bar{\mathcal{Y}} = m^{-1} \sum_{i=1}^{m} y^{(i)}$. The operator $\mathrm{diag}(\cdot)$ forms a diagonal matrix from the diagonal entries of Σ_T. $\sqrt{\mathrm{diag}(\Sigma_T)} \in \mathbb{R}^{N \times N}$ represents the surrogate error standard deviation at each observation location. This normalisation ensures that the scaling factor α_T is dimensionless, by relating the surrogate error to the magnitude of the observations. Where required, the model predictions are interpolated onto the observation locations such that consistent dimensions are obtained.

The purpose of Equations (6) to (8) is to combine the different sources of uncertainty, namely the surrogate error and the noise, consistently into the soil parameter calibration process. Incorporating these uncertainties increases the overall covariance of the error term, thereby enlarging the total uncertainty. This enlargement is beneficial for soil parameter calibration, as it makes the comparison between observations and FE model predictions less restrictive and consequently permits a wider range of plausible soil parameters. While the surrogate error covariance $\Sigma_T \in \mathbb{R}^{N \times N}$ can be computed using Equation (6), the noise covariance $\Sigma_{\varepsilon_r} \in \mathbb{R}^{N \times N}$ is typically unknown, and the two components are scale-sensitive, meaning they cannot be directly combined. Therefore, the consideration of Σ_T is incorporated via a proportional inflation factor $\alpha_T \in \mathbb{R}^{N \times N}$ in the noise covariance, preventing calibration dominated by a constant term Σ_T and moderately increasing uncertainty without altering the heteroscedastic structure, thereby improving stability. Assuming that $\Sigma_{\varepsilon_r T} \in \mathbb{R}^{N \times N}$ follows a zero mean multivariate Gaussian distribution, and given m sets of sparse independent observation vectors $\mathcal{Y}=(y^{(1)}, \cdots, y^{(m)})$, the likelihood can be expressed as:

$$\begin{aligned}\mathcal{L}(x,\varepsilon_r|\mathcal{Y}) &= \prod_{i=1}^{m} N(y^{(i)}|\tilde{\mathcal{M}}(x),\Sigma_{\varepsilon_r T}) \\ &= \prod_{i=1}^{m} \frac{1}{\sqrt{(2\pi)^N \det(\Sigma_{\varepsilon_r T})}} \exp\left(-\frac{1}{2}\left[y^{(i)}-\tilde{\mathcal{M}}(x)\right]^{\mathsf{T}} \Sigma_{\varepsilon_r T}^{-1}\left[y^{(i)}-\tilde{\mathcal{M}}(x)\right]\right),\end{aligned} \tag{9}$$

in which, $\det$ is the determinant. This choice is primarily for simplicity and mathematical convenience; alternative distributions could be adopted if supported by available data. Equation (9) now explicitly accounts for uncertainty associated with *forward parameters* x, *observation noise* ε_r, and *surrogate error* T. When the observations include multiple physical quantities (e.g., displacements, internal forces), the full observation set is organised into groups as $\mathcal{G} = \bigcup_{g=1}^{N_{gr}} \mathcal{G}^{(g)} = (\mathcal{G}^{(1)}, \cdots, \mathcal{G}^{(N_{gr})})$ with $\mathcal{G}^{(g)}{=}\mathcal{Y}{=}(y^{g(1)}, \cdots, y^{g(m)})$. The overall group $\mathcal{G}$ encompasses N_{gr} types of observations, and each $\mathcal{G}^{(g)}$ includes m sets of similar measurement conditions y. Assuming independence between the N_{gr} groups and the m sets of observations, the full likelihood function can be rewritten as:

$$\begin{aligned}\mathcal{L}(x,\varepsilon_r|\mathcal{G}) &= \prod_{g=1}^{N_{gr}} \mathcal{L}^{(g)}(x,\Sigma_{\varepsilon_r T}^{(g)}|\mathcal{G}^{(g)},\lambda^{(g)}) \\ &= \prod_{g=1}^{N_{gr}} \prod_{i=1}^{m} N(y^{g(i)},\lambda^{(g)}|\tilde{\mathcal{M}}(x),\Sigma_{\varepsilon_r T}^{(g)}).\end{aligned} \tag{10}$$

While *model discrepancies* are not represented directly, their influence is addressed via a *weighting* scheme using a predefined weight vector λ. Notably, this weighting scheme extends beyond discrepancy treatment and serves as a practical tool for emphasising regions deemed critical by engineering judgment. The regions of interest and their corresponding weights are predefined before calibration, representing areas of the problem where most of the stress and strain changes are expected, based on model responses relevant to the problem. For example, previous studies and applications have often focused on regions such as maximum and top wall deflections in excavation scenarios or pile-head displacements in offshore pile loading tests. The regions of interest may evolve as the analysis progresses. In staged excavations, for example, the location of maximum wall deflection may shift during construction. Each observation group $\mathcal{G}^{(g)}$ is assigned its own likelihood function $\mathcal{L}^{(g)}$ and corresponding weight vector $\lambda^{(g)}$, where, $y^{g(i)}$ denotes the i-th set of similar observations within the g-th group. As geotechnical construction and/or loading are typically staged, data are acquired incrementally, requiring sequential updates of model parameters (x, ε_r) through data assimilation. Assuming stagewise dependence, the sequence forms a *Markovian process*, i.e., the state at each stage depends only on the preceding stage (Bishop and Nasrabadi 2006), with recursive Bayesian updates, as illustrated in Figure 2. The horizontal arrows represent the propagation of the model parameters from one stage to the next, while the vertical arrows indicate the incorporation of observations at each stage. In this framework, the posterior distribution at stage $t-1$ is used as the prior for stage t, and the calibration at each stage is carried out using only the observations available at that stage. As ε_r is treated as stage-specific and re-estimated at each stage, it is not carried forward and is therefore excluded from the figure for clarity. This one-step propagation avoids conditioning on the full history, thereby reducing computational overhead and improving the efficiency of sequential calibration.

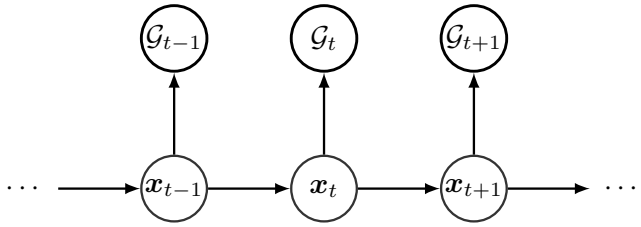


Figure 2: Sequential Bayesian inversion

To assess predictive capability and quantify uncertainty before and after each stage of calibration, both prior and posterior predictive distributions are computed, respectively. The former integrates over the prior $\pi(\mathrm{y}) = \int_{\mathcal{D}_{\boldsymbol{x}}} \pi(x)\pi(\mathrm{y}|x)\mathrm{d}x$, while the latter incorporates posterior knowledge via $\pi(\mathrm{y}|\mathcal{Y}) = \int_{\mathcal{D}_{\boldsymbol{x}|\mathcal{Y}}} \pi(x|\mathcal{Y})\pi(\mathrm{y}|x)\mathrm{d}x$. y denotes the model prediction in the output space, as provided by an appropriately constructed surrogate.

2.3 Reduced order model and global sensitivity analysis

Reduced-order models allow for flexible combinations of dimensionality reduction techniques and surrogate modelling approaches. In this study, *polynomial chaos expansion* (PCE) is adopted as the surrogate modelling technique. PCE represents the model response as an expansion of orthogonal polynomial basis functions of the uncertain input parameters. In this study, Legendre polynomials are adopted as the basis functions, with a hyperbolic truncation scheme and sparse regression used to identify the relevant polynomial terms (Wagner et al. 2020). Owing to the orthogonality of the polynomial basis, sensitivity indices can be computed analytically from the polynomial chaos expansion coefficients, without requiring additional Monte Carlo sampling (Sudret 2008; Wagner et al. 2020; Nagel et al. 2020; Yang et al. 2026).

Global sensitivity analysis is conducted via Sobol' indices, which decompose the output variance into contributions from individual inputs and their interactions. Based on the orthogonal decomposition, the total variance of the model output $\mathcal{M}(x)$ is defined as $D_T := \mathrm{Var}[\mathcal{M}(x)]$, and the total Sobol' indices are calculated as:

$$S_i^T := \sum_{u \supset i} S_u = \sum_{u \supset i} \frac{\mathrm{Var}[\mathcal{M}_u(x_u)]}{D_T}, \tag{11}$$

where the terms $\mathcal{M}_u(x_u)$, with $u := i_1, \cdots, i_s \subset 1, \cdots, M$, represent the decomposition based on subsets of input variables. Further details on the reduced-order model construction and the associated sensitivity analysis are provided in Yang et al. (2026).

When the model response is vector-valued, such as spatially distributed deflections at multiple depths, constructing a polynomial chaos expansion surrogate for each output component can become computationally demanding. This is particularly relevant when the number of output locations is large. To address this, principal component analysis (PCA) is employed to reduce the dimensionality of the model outputs prior to surrogate construction. The original response is projected onto a reduced set of dominant modes, and polynomial chaos expansion is then constructed in this

reduced space. This principal component analysis-polynomial chaos expansion combination reduces the number of surrogate outputs while preserving the dominant correlation structure of the vector-valued response (Yang et al. 2026). It is noted that several alternative reduced-order model configurations were also explored by the authors, including autoencoders, singular value decomposition combined with neural networks, tree-based models, and Gaussian process surrogates. None of these approaches, however, exhibited consistently superior performance; the relative performance of different surrogate modelling techniques may depend on the size and coverage of the training data.

2.4 Custom likelihood function

Model calibration is performed to infer the unknown parameters (x, ε_r) using the observation groups $\mathcal{G}$, as formulated in Equations (9) and (10). The *surrogate error* and *model discrepancy*, being known, are not treated probabilistically. In many geotechnical applications, the noise covariance Σ_{ε_r} is simplified as $\sigma^2 \times \mathbf{I}$, where $\sigma > 0$ is an unknown scalar noise level and $\mathbf{I} \in \mathbb{R}^{N \times N}$ denotes the identity matrix, implying independent and identically distributed noise (Hsein Juang et al. 2013; Kelly and Huang 2015; Qi and Zhou 2017; Lo and Leung 2019; Jin et al. 2021; Asem and Gardoni 2021). While computationally convenient, this assumption may not reflect the real site conditions. For instance, instruments such as inclinometers may produce spatially aggregated responses with noise levels varying with depth, leading to heteroscedasticity or spatial correlations. A more general representation is proposed by assuming that the varying noise level σ can be approximated using low-order polynomial basis functions. A diagonal form is adopted as a practical simplification, which avoids the need to estimate a full covariance matrix and improves the numerical stability of the inversion process, particularly when the number of observations is large. This ensures smoothness and reduce dimensionality of the noise term. The noise covariance matrix can be written as:

$$\Sigma_{\varepsilon_r} = \mathrm{diag}\big((\sigma_1^2, \ldots, \sigma_N^2)\big), \tag{12}$$

$$\sigma(\omega) = (\sigma_1, ..., \sigma_N)^{\mathsf{T}} = \sum_{k=0}^{p} \omega_k \times \psi_k \times t, \tag{13}$$

where ψ_k is the k-th Legendre polynomial basis, ω_k is the associated coefficient, and p is the polynomial order. The polynomial is defined along the pile length, with the depth coordinate scaled between the top and bottom of the pile, allowing the noise level to vary continuously along the pile profile. The scaling factor $t = \max(\mathcal{Y})$ is introduced to dynamically link the noise level to each construction/loading stage, thereby capturing the stage-dependent variability in noise. Since this varying noise level is typically unknown, it is integrated and calibrated within the noise covariance matrix. The direct parametrisation may correlate with the FE mesh nodes or the sensor distribution, and may result in a prohibitively large parameter space. To achieve a substantial reduction in this parameter set, the derivation relies on employing a continuous polynomial basis. Utilising different orders of Legendre polynomial basis enables the effective approximation of the arbitrary noise level, thereby condensing the whole noise parametrisation into a manageable set of polynomial coefficients. This hierarchical formulation follows $\Sigma_{\varepsilon_r} \in \mathbb{R}^{N \times N} \rightarrow \sigma \in \mathbb{R}^{N} \rightarrow \omega \in \mathbb{R}^{p+1} := (\omega_0, \cdots, \omega_p)$. This provides a compact parametrisation of the noise model while preserving variability in the noise model (Nagel et al. 2020; Wagner et al. 2020). This approach significantly reduces the complexity of inferring the

full covariance matrix, transforming a high-dimensional problem into a lower-dimensional one with a limited, well-constrained set of hyper-parameters. The hyper-parameter ω in our geotechnical context represents the confidence or tolerance of the observed data. A low value of ω indicates high confidence in the data, suggesting that the observations are likely accurate. Conversely, a high value implies greater uncertainty in the observations. The complete set of calibration parameters is (x, ω), where x denotes the uncertain soil parameters.

The custom likelihood formulation and the associated parameter dependencies are summarised in Algorithm 1. It is acknowledged that different representations of the noise covariance can influence the calibration process. Common approaches include assuming identical and independent noise covariance (homoscedasticity), using a full covariance matrix to account for spatial correlations (heteroscedasticity) and varying noise levels, or considering only varying noise levels with no spatial correlation (diagonal heteroscedastic matrix). In some cases, customized noise covariance matrices may be developed based on the specific data collection methods employed. However, these representations are beyond the scope of this study.

Algorithm 1: Custom-defined log-likelihood computation for multiple-source observations

Input: Parameters of interest: (x, ω);
Observation groups: $\mathcal{G} = \bigcup_{g=1}^{N_{\text{gr}}} \mathcal{G}^{(g)}$ with $\mathcal{G}^{(g)} = \{y^{g(1)}, \ldots, y^{g(m)}\}$;
Surrogate error representation: T;
Emulator: $\tilde{\mathcal{M}}$;
Number of Markov Chain Monte Carlo chains: N_{chain};
Weight vector: λ

Output: Log-likelihood sum vector: $[\log \mathcal{L}]$

Set $m \leftarrow$ number of similar measurements in each group;
Set $N_{\text{gr}} \leftarrow$ number of observation groups;
Initialize $\log \mathcal{L}_j \leftarrow 0$, for $j = 1, \ldots, N_{\text{chain}}$;
for $j = 1$ **to** N_{chain} **do**
  for $g = 1$ **to** N_{gr} **do**
    for $i = 1$ **to** m **do**
      Evaluate $\Sigma^{(g)}_{\varepsilon_r T}$ using $(\omega^{(g)}, T^{(g)})$ via Equations (6) and (7);
      Compute log-likelihood increment:

$$\Delta \log \mathcal{L}_j^{(g,i)} = \log \left(\mathcal{N} \left(y^{g(i)}, \lambda^{(g)} \middle| \tilde{\mathcal{M}}(x), \Sigma^{(g)}_{\varepsilon_r T} \right) \right),$$

      Update cumulative sum:

$$\log \mathcal{L}_j \leftarrow \log \mathcal{L}_j + \Delta \log \mathcal{L}_j^{(g,i)},$$

    end for
  end for
  Normalize:

$$\log \mathcal{L}_j \leftarrow \frac{1}{N_{\text{gr}}} \times \frac{1}{m} \times \log \mathcal{L}_j,$$

end for
return $[\log \mathcal{L}_1, \ldots, \log \mathcal{L}_{N_{\text{chain}}}]$

2.5 Bayesian workflow

The complete inference structure is illustrated in the generalized probabilistic framework shown in Figure 3. This framework integrates prior engineering knowledge on both soil parameters and noise hyper-parameters, enabling physically-informed Bayesian inference. Observations are collected across multiple physical groups and stages. For each group, the m observations in Equation (10) denote repeated measurements of the same response quantity at a given stage. To account for various sources of uncertainty in Section 2.1, including *model discrepancy*, *surrogate error*, and *observational noise*, an extended Kennedy and O'Hagan (2001) structure is adopted. Observations, subject to noise, are first collected at each stage and compared with the computational (FE) model to identify regions of critical interest and potential model discrepancy. The surrogate is constructed over the admissible parameter ranges to ensure adequate coverage of the parameter space, while the prior distributions define the probabilistic assumptions in the subsequent Bayesian calibration and may differ from the ranges used for surrogate construction. *Surrogate error* is quantified based on the available test dataset. Prior distributions are assigned to the soil parameters uncertainties, and posterior inference is performed via Markov chain Monte Carlo sampling, with all sources of uncertainty incorporated within the custom-defined log-likelihood framework. A spatial weighting strategy is used within to enhance the engineering significance and avoid the potential effect from *model discrepancy*. All components are systematically embedded within a custom-defined likelihood function that reflects both the hierarchical structure of the data and the uncertainty sources. The inference process is carried out using a sequential Bayesian calibration strategy, with posterior samples drawn via Markov chain Monte Carlo and subsequently used for predictive assessment and uncertainty quantification.

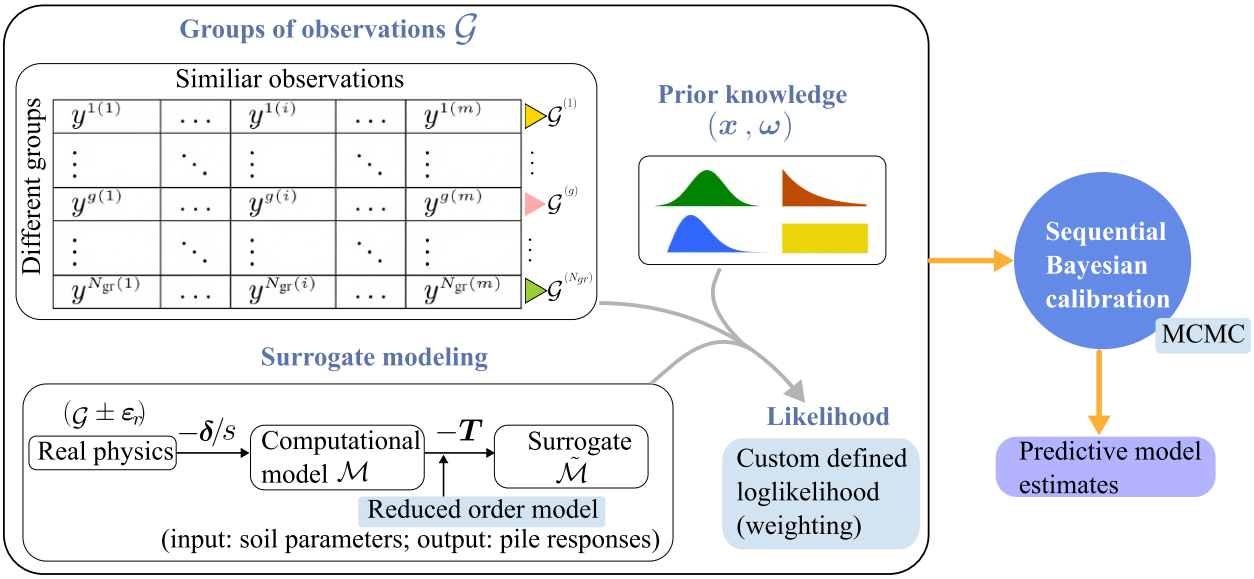


Figure 3: A generalized framework

3 Application-PISA pile problem

3.1 Problem setup

To assess the performance of the proposed probabilistic framework, a laterally loaded monopile problem is examined, based on field and numerical investigations conducted within the PISA project for offshore wind turbine (OWT) foundations (Byrne et al. 2020; Zdravkovic et al. 2020b). The study presents results for a specifically selected monotonic quasi-static loaded pile, incorporating high-quality field and laboratory data from the Cowden site in the UK (Zdravkovic et al. 2020a). Numerical simulations were conducted using the *Imperial College Finite Element Program*, ICFEP, (Potts and Zdravkovic 1999). Model predictions were validated against field measurements by Zdravkovic et al. (2020b), with a focus on two response quantities: (i) lateral pile deflections below the mudline and (ii) horizontal force at maximum eccentricity above mudline versus the mudline horizontal displacement, as illustrated in Figure 4. The pile considered conforms to the CL2 type defined by Zdravkovic et al. (2020b), with a diameter D = 2.0 m and an embedded length-to-diameter ratio L/D = 5.25. Monotonic lateral loading was applied at a height of 10 m above the ground surface (see the schematic in Figure 4). The glacial clay was represented using an extended modified Cam Clay formulation, augmented by a non-linear Hvorslev surface and a non-linear small-strain overlay. The open-ended pile was discretised with eight-noded shell elements (Schroeder et al. 2007). Soil-pile interaction was modelled using 16-noded zero-thickness interface elements (Day and Potts 1994), governed by an elasto-plastic Tresca law with elastic normal and shear stiffnesses, and zero tensile capacity. The interface shear strength in compression was set equal to the initial undrained shear strength of the adjacent soil elements, thereby capturing spatial variations associated with the in-situ stress state, overconsolidation ratio and loading direction. For more details related to the pile geometry, soil constitutive models, and boundary conditions, refer to Zdravkovic et al. (2020b). A total of nine incremental horizontal loading stages were applied at the pile head. For two selected stages, corresponding to lateral displacements of v_G = 20 mm (stage 4) and v_G = 200 mm (stage 9), pile deflections were recorded at 101 positions down the embedded shaft, forming the first observational dataset: $\mathcal{G}^{(1)} = \{y^{(i)}\}_{i=4,9},\ y^{(i)} \in \mathbb{R}^{101}$. The second dataset comprises the horizontal applied force at pile top at various loading stages, denoted by $\mathcal{G}^{(2)} = \{y^{(i)}\}_{i=1}^{9},\ y^{(i)} \in \mathbb{R}^{1}$, where $i = 1, \ldots, 9$ denotes the loading stages.

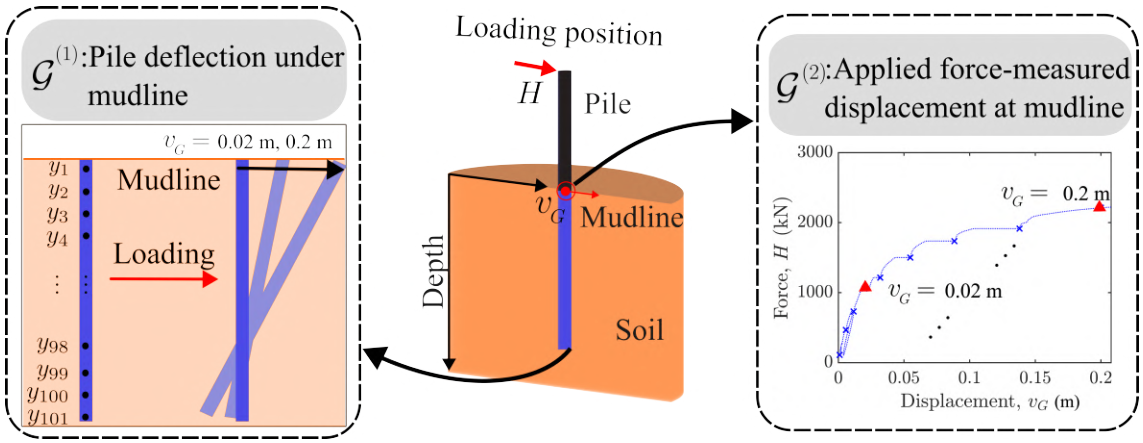


Figure 4: Model setup for a laterally loaded pile: $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$

The uncertain soil parameters x comprises: the initial small-strain shear modulus G_0, the coefficient of earth pressure at rest K_0, and the overconsolidation ratio OCR (see Table 1). These parameters were identified, through the experience gained during the PISA project, as those (i) exhibiting the greatest uncertainty in their interpretation from the available site and laboratory data; (ii) contributing directly to the derivation of the undrained shear strength S_u adopted for design; and (iii) initialise the maximum shear stiffness (G_0) and stress state (K_0 and OCR) in the soil. It is noted that the three soil parameters considered in this study are taken to vary with depth. Accordingly, the quantities listed in Table 1 are not specified directly as constitutive model parameters at each depth, but are treated as profile-defining parameters that control the depth-dependent soil property profiles assigned to the FE model. For example, as shown in Figure 5, the stiffness-related parameter (e.g. G_0) controls the gradient of the stiffness profile, while parameters such as K_0 and OCR shift the corresponding profiles by modifying the value at ground level, without altering their assumed shape. Once these profiles are defined, the corresponding values at each depth are assigned to the soil layers and used in the FE analysis. This provides a practical link between the calibrated geotechnical quantities and the parameters used in the constitutive model, while avoiding the need to directly calibrate parameters at every depth. For simplicity of notation, these quantities are still referred to as "soil parameters" throughout this paper. These soil parameters were modelled as independent random variables drawn from a prior distribution $\pi(x)$ defined over the domain $\mathcal{D}_{\boldsymbol{x}}$. Uniform prior distributions were adopted over physically credible ranges derived from the available site investigation and laboratory data. The input variables are assumed to be independent, which is compatible with the PCE surrogate construction adopted in this study. The upper and lower limits adopted are derived from the ground characterisation of the Cowden site as documented by Zdravkovic et al. (2020b) and Yang et al. (2026). All other model definitions and settings follow the calibrated PISA reference models of Zdravkovic et al. (2020b) and are retained to provide a consistent baseline. The sequence of data acquisition and model updating is depicted in Figure 6, where subscripts denote loading stage and superscripts indicate observation group. As pile deflections were only recorded at stages 4 and 9, data at other stages were treated as missing in the deflection channel. Consequently, applied loads were used for model calibration at all construction stages, while deflection profiles were additionally considered at stages 4 and 9.

Table 1: Priors for parameter variations

x	$\mathcal{D}_{\boldsymbol{x}}$	Unit	Physical meaning
G_0	$\mathcal{U}$(55000,165000)	kPa	Small-strain shear modulus (adopted range of p'-dependent values of G_0)
K_0	$\mathcal{U}$(1.2,1.8)	–	Adopted range of values for at-rest coefficient of earth pressure at ground level
OCR	$\mathcal{U}$(15,50)	–	Adopted range of values for overconsolidation ratio at ground level

3.2 Performance of surrogate and global sensitivity analysis

Latin hypercube sampling was employed, and models were incrementally executed until 36 runs were found to provide satisfactory surrogate performance according to the error metrics defined below (less than 3%). All FE model runs were carried out (over six hours per run) on a 32-core Intel® Xeon® Silver 4216 CPU (2.10 GHz) with 256 GB RAM. Of these, $K = 31$ simulations were used for training the surrogate, with the remaining $K' = 5$ for test validation.

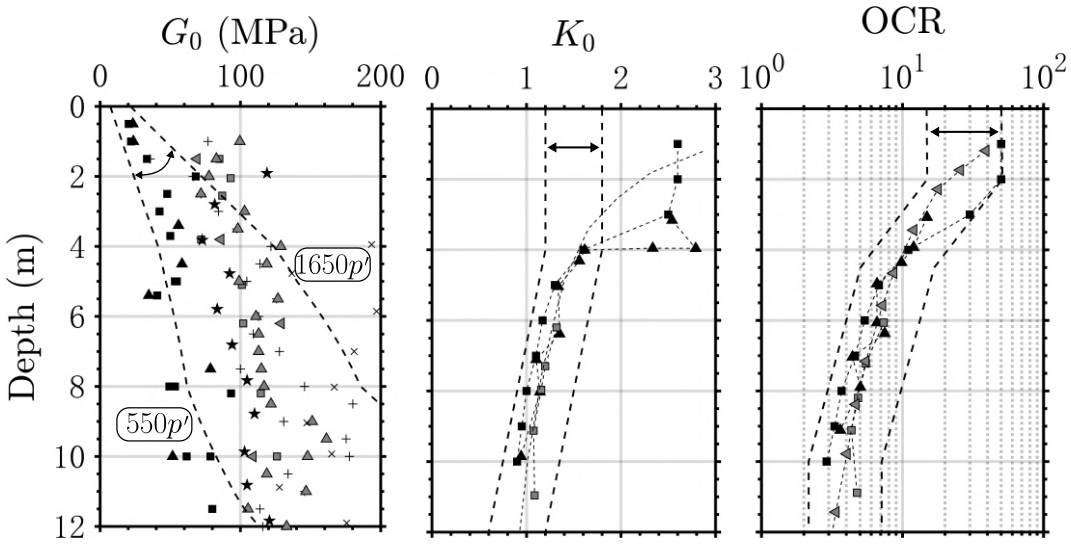


Figure 5: Soil uncertainties

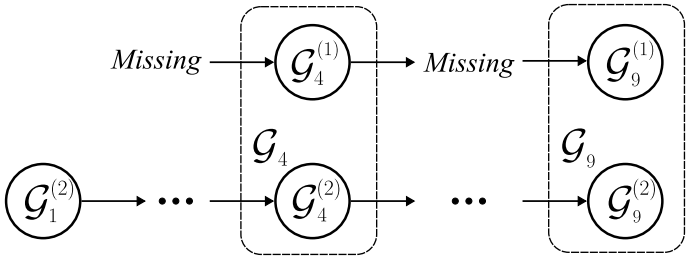


Figure 6: Data collection sequence

The number of 36 was chosen because the input space (G_0, K_0 and OCR) is relatively low-dimensional and exhibits near-linear behaviour, which facilitates interpolation and reduces the number of required evaluations. In addition, the output fields (deflection and applied force) were projected onto a low-dimensional latent space using *principal component analysis*, where only a small number of principal components were needed to represent the dominant features of the high-dimensional response. A reduced order model based on *principal component analysis-polynomial chaos expansion* was then developed for the vector-valued deflections $\mathcal{G}^{(1)}$ and the scalar-valued forces $\mathcal{G}^{(2)}$. For $\mathcal{G}^{(1)}$, the dimensionality was reduced by selecting the number of principal components N' to retain 99.99% of the total output variance, yielding $N' = 1$. This result demonstrates that the high-dimensional response $y \in \mathbb{R}^{101}$ can be effectively captured by a single dominant mode, reflecting the smooth and coherent nature of the deflection patterns. The combination of *principal component analysis* with *polynomial chaos expansion* offers several advantages, including (1) improved model prediction and the ability to capture correlations among outputs, (2) analytical global sensitivity analysis, (3) enhanced accuracy for subsequent calibration, (4) less requirements for the high number of FE runs and (5) accelerated calibration, as it is based on the subspace rather than the entire dataset. A detailed investigation is provided in Yang et al. (2026).

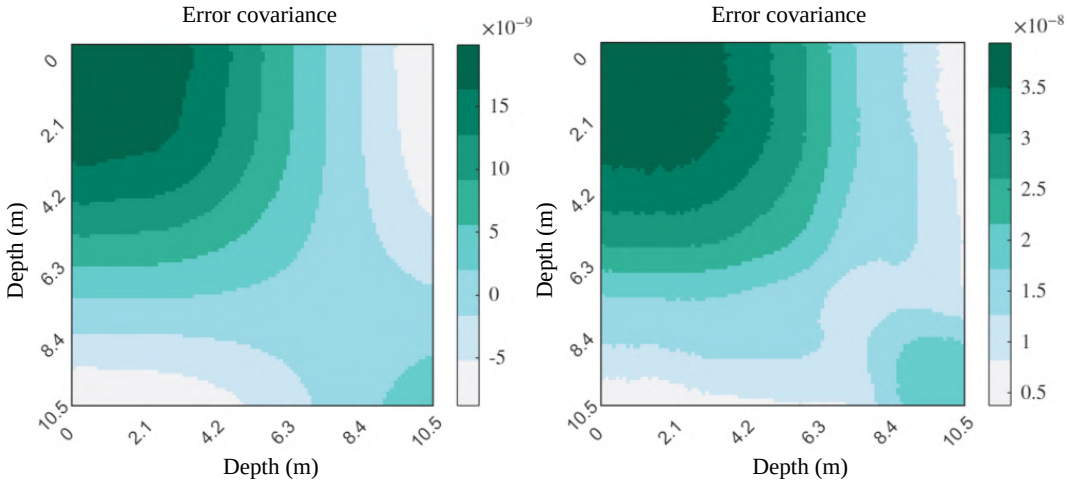
The performance of the surrogate models for $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$ is evaluated using the *Global Normalized Absolute Error* (GNAE), defined as:

$$\text{GNAE} = \frac{1}{K'} \sum_{k=1}^{K'} \frac{\left|\mathcal{M}(x_k) - \tilde{\mathcal{M}}(x_k)\right|}{\max(\mathcal{M}(x_k)) - \min(\mathcal{M}(x_k))} \times 100. \tag{14}$$

The corresponding *Global Normalized Absolute Error* for each stage is summarized in Table 2. A single value cannot fully capture the surrogate error in vector outputs for $\mathcal{G}^{(1)}$. As outlined in Equation (6), Σ_T is computed as shown in Figure 7 and incorporated into the likelihood function. The two panels show the surrogate error covariance for nodal displacements at (a) Stage 4 and (b) Stage 9. Depth are measured from the mudline to the pile base, as shown in Figure 4; the "ends" therefore refer to the near-mudline and near-toe regions, and the transition region in the middle where displacements are small and tend to zero. The error is accordingly low in this transition region and higher where displacements are mobilised. Furthermore, the measured displacement at stage 9 (v_G = 200 mm) is larger than at stage 4 (v_G = 20 mm), which results in a larger surrogate error. Since the surrogate error is expressed in terms of absolute displacement, the larger movement at stage 9 is likely to be associated with greater error magnitudes, although the relative accuracy of the surrogate model appears broadly comparable at the two stages.

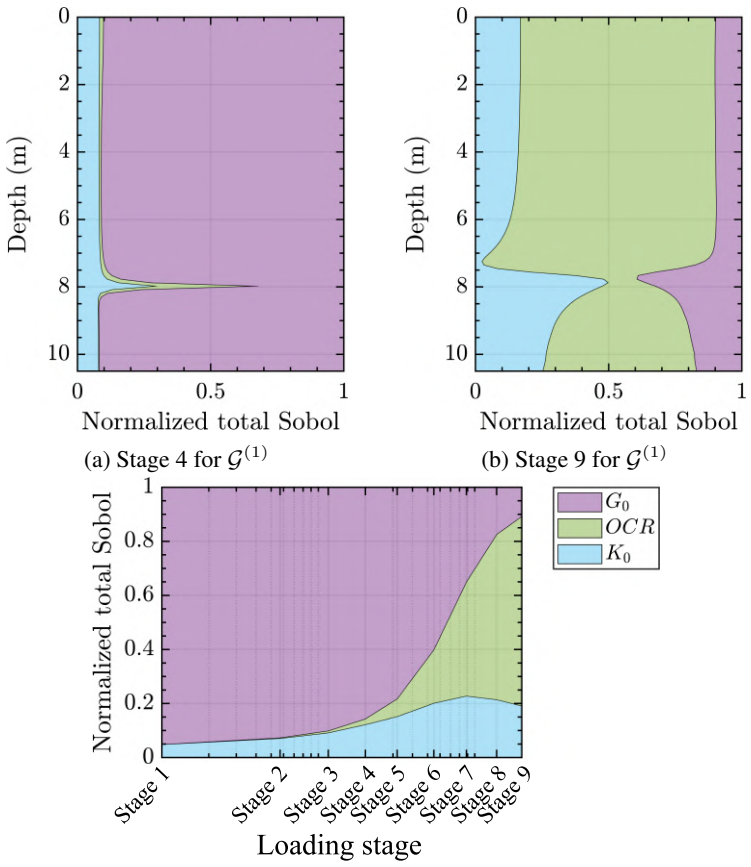
Table 2: Surrogate error and covariance for each loading stage

Stage	GNAE ($\mathcal{G}^{(1)}$)	GNAE ($\mathcal{G}^{(2)}$)	Σ_T ($\mathcal{G}^{(1)}$) (m)^2	Σ_T ($\mathcal{G}^{(2)}$) (kN)^2
1	–	0.56%	–	0.01
2	–	2.29%	–	1.58
3	–	2.79%	–	8.22
4	2.23%	2.49%	Figure 7(a)	16.11
5	–	2.54%	–	31.06
6	–	2.91%	–	71.28
7	–	1.06%	–	9.72
8	–	2.08%	–	50.05
9	1.40%	0.90%	Figure 7(b)	11.91



(a) Stage 4 (b) Stage 9

Figure 7: Surrogate error covariance matrix Σ_T



(a) Stage 4 for $\mathcal{G}^{(1)}$

(b) Stage 9 for $\mathcal{G}^{(1)}$

(c) All stages for $\mathcal{G}^{(2)}$

Figure 8: Sobol's indices

A global sensitivity analysis was performed to ensure that the parameters under investigation show a non-trivial influence on the model outputs. Sobol' indices computed for both output types across all loading stages are presented in Figure 8. The results show that the three parameters have a primary influence on the outputs. Sobol' indices in Figure 8(a) indicate that, at Stage 4 loading ($v_G = 20$ mm), the soil response is dominated by the shear stiffness parameter G_0. K_0 also contributes additionally by controlling the ground horizontal effective stress, and therefore the stress level governing G_0. This indicates the soil response at early stages of loading and Figure 8(c) confirms that the Stage 4 is the boundary to working load conditions governed by soil stiffness. Stage 9 Sobol' indices in Figure 8(b) demonstrate the dominance of OCR and K_0 that influence the soil's undrained shear strength at later stages of loading that lead to plastic failure in the soil. Figure 8(c) demonstrates that the increasing influence of these two parameters post-Stage 4 loading. It is noted that an abrupt change occurs at depth of 8 m, where the pile deflection near this region approaches zero. This coincides with the point of rotation of a rigid pile (low L/D ratio) in the overall mechanism of pile response to lateral loading. This is a typical transition region where values near zero lead to value instability. In regions where sensitivity is not clearly manifested or weakly dependent on model parameters, relatively small weights may be appropriate. Although this transition region was not explicitly emphasised in the present study due to the

negligible displacement magnitude, it nevertheless indicates a possible refinement of weighting strategies through the incorporation of sensitivity information.

3.3 Custom likelihood and noise parametrization

In addition to the forward soil parameters listed in Table 1 and the known surrogate error covariance Σ_T shown in Figure 7, the observation noise ε_r should also be explicitly parametrised. For the vector-valued pile deflections, both the magnitude and structure of noise may vary with depth. Depth-dependent noise levels $\sigma = \{\sigma_1, \cdots, \sigma_{101}\}$ are approximated by a third-degree Legendre polynomial ($p = 3$), as defined in Equation (15), which is assumed to be sufficient to characterise the noise:

$$\sigma(z_i) = \sum_{k=0}^{3} \omega_k^{(1)} \times \psi_k(2 \times \frac{z_i - z_1}{z_{101} - z_1} - 1) \times \max(\mathcal{G}_t^{(1)}), \tag{15}$$

in which, z_i denotes the pile depth coordinate with 101 points. For scalar force outputs $\mathcal{G}^{(2)}$, the observation noise ε_r is characterised by a single standard deviation, defined as $\sigma = \omega_0^{(2)} \times \max(\mathcal{G}_t^{(2)})$, without involving a covariance matrix. All hyper-parameters are dimensionless and scaled by the output magnitude $\mathcal{G}_t$ at each stage, ensuring consistency with the observational scale. For $\mathcal{G}^{(1)}$, the noise covariance is controlled via four parameters $(\omega_0^{(1)}, \omega_1^{(1)}, \omega_2^{(1)}, \omega_3^{(1)})$ that define the polynomial approximation, while $\mathcal{G}^{(2)}$ is governed solely by $\omega_0^{(2)}$. Uniform priors are assigned, as listed in Table 3. To introduce judgment on $\mathcal{G}^{(1)}$, particular weighting emphasis is placed on the upper pile portion (2 m below the mudline, as this region experience the largest changes in stresses and strains due to the lateral loading of the pile). Accordingly, a predefined weight λ of 5.0 is applied in this region to enhance its influence during calibration, while a baseline weight of 1.0 is maintained elsewhere. It is noted here that the weight factors are prescribed by engineers according to problem-specific considerations. Their absolute magnitudes are not of consequence, as they are not sensitive within the likelihood space. The calibration is influenced only by the relative contrast between critical and non-critical regions, which reflects their engineering judgment and may vary among boundary value problems. The critical-zone depth and the value of λ are treated here as operational, problem-specific modelling choices rather than uniquely prescribed or formally optimised quantities. Guided by response sensitivity and exploratory trial calibrations, increased weight is applied to a representative near-surface zone and several (depth, λ) combinations are examined. The configuration (top 2 m, λ = 5) is adopted for the present study as it gives stable calibration behaviour and acceptable agreement with the observations while maintaining the intended emphasis on the critical region. The selected values should not be taken as general guidance; further work is required to develop systematic procedures for defining weighted regions and selecting associated weighting factors. In large-scale monitoring, FE model outputs and field data are often collected in matrix or tensor formats. When encountering such multidimensional data, weights can be assigned by flattening these structures into vectors, thereby ensuring flexibility in handling complex data while maintaining focus on key calibration parameters. The weight vector also needs to match the dimensionality of the observations.

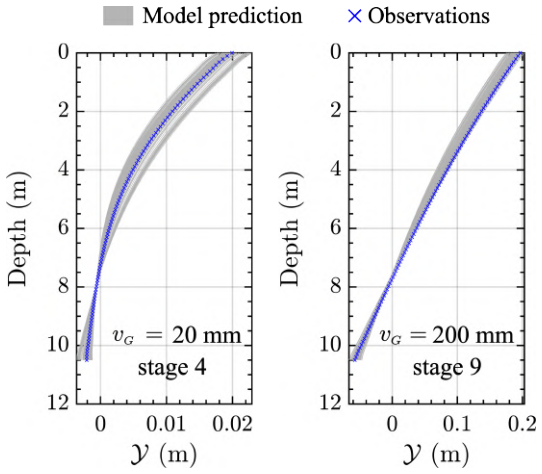
Table 3: Priors for noise hyper-parameters and unit for $\mathcal{G}^1$ and $\mathcal{G}^2$

ω	$\mathcal{D}_{\boldsymbol{\omega}}$	Unit	Physical meaning
$\omega_0^{(1)}$	$\mathcal{U}$(-0.1,0.1)	–	Hyper-parameter for $\mathcal{G}^{(1)}$
$\omega_1^{(1)}$	$\mathcal{U}$(-0.1,0.1)	–	Hyper-parameter for $\mathcal{G}^{(1)}$
$\omega_2^{(1)}$	$\mathcal{U}$(-0.1,0.1)	–	Hyper-parameter for $\mathcal{G}^{(1)}$
$\omega_3^{(1)}$	$\mathcal{U}$(-0.1,0.1)	–	Hyper-parameter for $\mathcal{G}^{(1)}$
$\omega_0^{(2)}$	$\mathcal{U}$(0,0.05)	–	Hyper-parameter for $\mathcal{G}^{(2)}$

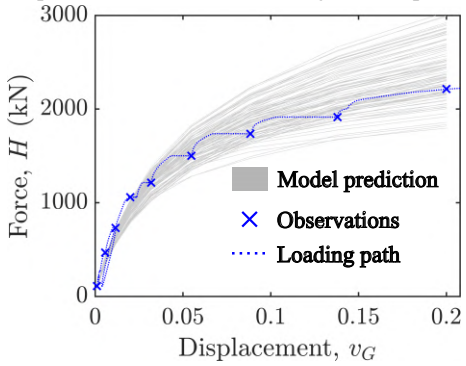
3.4 Inversion results

A prior predictive check is first conducted to examine whether the FE model is capable of reproducing the range of observed responses. This step provides an initial assessment of model fidelity and highlights any potential sources of discrepancy. This is essential in geotechnical engineering, where establishing the consistency between numerical predictions and field measurements forms the basis for building confidence in, and guiding, subsequent analyses. Figure 9 presents the prior predictive responses of the FE model for $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$, showing generally good agreement with the field observations and indicating that the model provides a credible representation of the pile behaviour. The gray curves correspond to prior FE predictions generated from the soil parameter uncertainties listed in Table 1, while the blue curves denote the measured responses. The stage 4 ($v_G = 20$ mm) and stage 9 ($v_G = 200$ mm) deflection profiles in Figure 9(a) are typical of laterally loaded piles, with peak displacement at the mudline and a transition to negligible lateral movement at a depth of pile rotation, which is at around 8 m. Figure 9(b) shows the corresponding applied force at the pile head, where the gray curves represent the FE predictions, the blue dashed curves denote the raw field measurements, and the blue crosses indicate the selected observations incorporated in the calibration. Posterior inferences are conducted via Markov Chain Monte Carlo sampling within a custom likelihood framework, using 100 parallel chains and 400 iterations per chain, with a burn-in fraction of 70%. To prove the effect of multiple-source observation fusion within the probabilistic method in this study, three cases are considered: (i) Case A: using only deflection observation $\mathcal{G}^{(1)}$; (ii) Case B: using only force observation $\mathcal{G}^{(2)}$; (iii) Case C: using both data types $\mathcal{G}^{(1)} + \mathcal{G}^{(2)}$;

Following sequential Bayesian calibration, samples drawn from the posterior distributions of K_0, G_0, and OCR are evaluated through the trained surrogate model to generate the corresponding pile responses. Figure 10 presents the resulting posterior predictive estimates for all three calibration cases. As indicated in Figure 6, pile deflection observations are available only at stages 4 and 9, while force observations exist for all loading stages. For simplicity, only the results for stages 4 and 9 are shown. Figure 10(a) shows the inferred soil uncertainties on the predicted pile deflection. The posterior predictive envelopes (gray regions) and MAP estimates (maximum a posterior, which corresponds to the most probable value) show good consistency with the measured profiles (blue lines). Figure 10(b) shows the corresponding posterior predictions of the applied force at the pile head. The results clearly illustrate the progression of reduction in uncertainty from stage 4 to stage 9. At stage 4, the uncertainty in the predicted applied force remains relatively large, and the maximum a posterior estimate shows noticeable deviation from the field measurements, reflecting the limited observational information available at this early stage. As the loading advances



(a) $\mathcal{G}^{(1)}$: pile deflection observations against FE predictions



(b) $\mathcal{G}^{(2)}$: measured mudline displacements and pile head loading against FE predictions

Figure 9: Prior predictive estimates

to stage 9 and additional observations are incorporated, the maximum a posterior estimate aligns more closely with the measured responses, and the posterior predictive bands provide a substantially improved representation of the recorded behaviour.

In Case A, where only pile deflections are used for calibration, the posterior distributions of the soil parameters exhibit a clear uncertainty shrinking (which can be seen from the samples in Figure 11(a) and Figure 11(d)), resulting in reduced predictive uncertainty in the corresponding deflection profiles (Figure 10(a)). A similar trend is observed in Case B, where only the applied forces are assimilated, as reflected in the posterior contraction shown in Figure 10(b), Figure 11(b), and Figure 11(e). Although both datasets originate from the same project and correspond to identical subsurface conditions, the independent use of $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$ produces markedly different posterior parameter distributions (see Figure 11(d) and Figure 11(e)), despite each case achieving good agreement with its respective observational quantity. This divergence indicates that soil parameters calibrated from one type of response, such as deflections in Case A, are unlikely to reproduce the alternative response, such as the applied forces in Case B, and vice versa.

In Case C, both datasets are assimilated jointly within the custom likelihood framework. The resulting posterior predictive estimates (Figure 10(c)) exhibit consistently good agreement with the observations for both $\mathcal{G}^{(1)}$ and

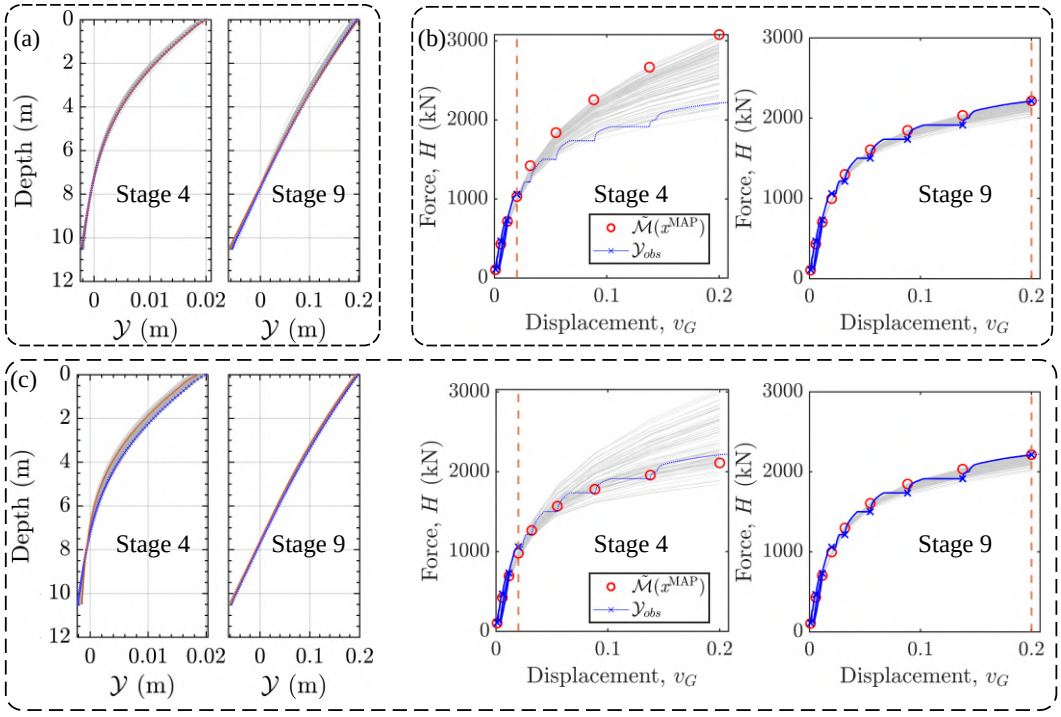


Figure 10: Posterior predictive estimates: (a) Case A (only pile displacement); (b) Case B (only applied force); (c) Case C (pile displacement and applied force)

$\mathcal{G}^{(2)}$, and the corresponding posterior distributions in Figure 11(c) and (f) show improved coherence. This behaviour illustrates a key advantage of Bayesian inference: integrating multiple, conditionally independent sources of information enhances parameter identifiability and more effectively constrains uncertainty by leveraging complementary sensitivities through the multiplicative nature of the likelihood. As shown in Figure 11, the inferred soil parameters in Case C converge to intermediate values relative to those obtained in Cases A and B. Each observation group, $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$, favours different regions of the parameter space, and their joint contribution shifts the posterior mode toward a compromise that satisfies both. Although the posterior in Case C is slightly broader, it provides a more faithful representation of the epistemic uncertainty supported by the complete dataset. The sequential uncertainty quantification process, in which stage-wise posterior updates reflect both the accumulation of observations and the evolving parameter identifiability, is further illustrated in the supplementary figures provided in the accompanying code.

Finally, it is noted that agreement between field observations and FE model predictions does not, by itself, ensure that the calibrated soil parameters are physically meaningful. In geotechnical practice, soil parameters obtained through optimisation related methods are often found to be unrealistic or inconsistent with site investigation data. Such outcomes are unacceptable, as reliable soil characterisation underpins all geotechnical analyses. As such, it is essential to validate the calibrated parameters against the original site investigation profiles along the depth to ensure their consistency with both the geotechnical conditions and the specific project conditions.

In Figure 12(a), the Cowden site conditions reported by Zdravkovic et al. (2020a) and Zdravkovic et al. (2020b) are compared with the soil parameters derived in this study. Four parameters profiles are involved: G_0, K_0, OCR,

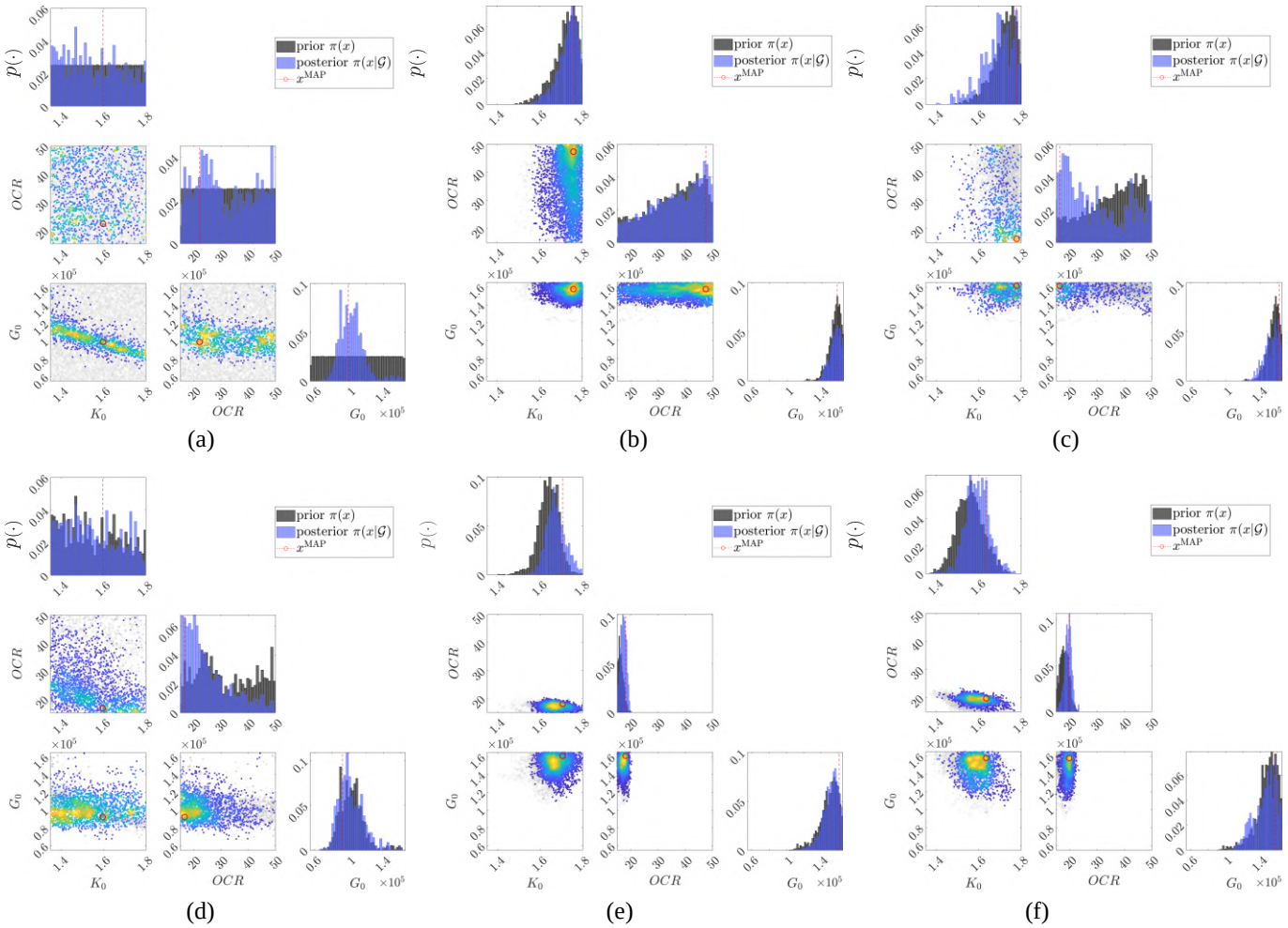
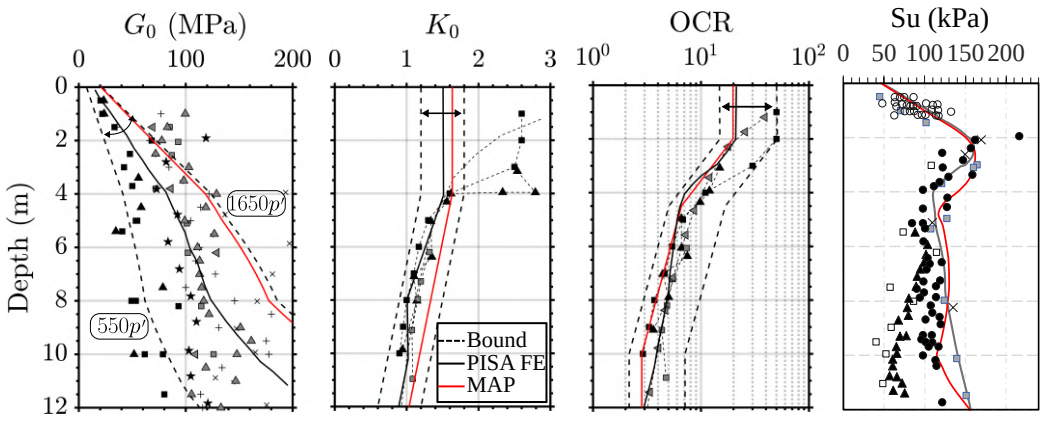
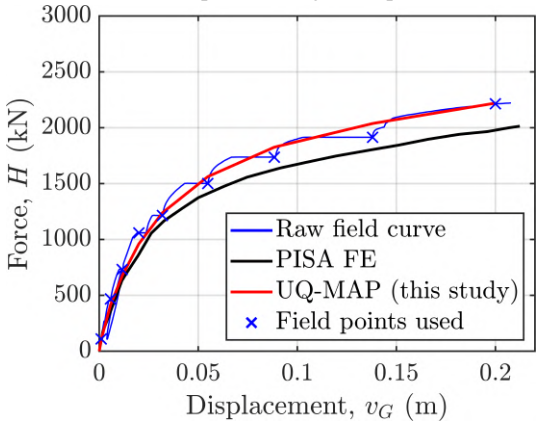


Figure 11: Corner plots: (a) CA at S4; (b) CB at S4; (c) CC at S4; (d) CA at S9; (e) CB at S9; (f) CC at S9; Note: S-Stage, CA-Case A, CB-Case B, CC-Case C

and the undrained shear strength S_u, the latter being derived. Detailed derivations can be found in Zdravkovic et al. (2020b). All parameter profiles are plotted along the depth, with gray dashed lines indicating the bound limits used in Table 1, gray solid lines representing the best fitted curve derived deterministically for site data from Zdravkovic et al. (2020b), and red lines denoting the most likely soil parameters x^{MAP} obtained through the uncertainty quantification in the current study. The calibrated parameters show good agreement with the reported site data, and the reconstructed undrained shear strength is also reasonably consistent with the Cowden site measurements and predicted soil profile used in Zdravkovic et al. (2020b). Furthermore, Figure 12(b) compares the observed field measurements, the predicted pile applied force response based the x^{MAP} in this study, and the FE predictions reported by Zdravkovic et al. (2020b). The satisfactory agreement among these datasets provides further confidence that the calibrated soil parameters are physically meaningful and consistent with the site conditions.



(a) Soil profile along the depth



(b) Applied force

Figure 12: Verification for the soil parameter obtained against results from Zdravkovic et al. (2020b)

4 Conclusion

This work enhances the sequential Bayesian inversion framework to address key challenges related to geotechnical problems, demonstrated through a full-scale laterally loaded pile test, leading to the following engineering conclusions:

(i) The reduced order model surrogate used captures spatial correlations in geotechnical responses, providing reliable modelling. The *principal component analysis-polynomial chaos expansion* based reduced-order model accurately captures the deflection profiles along the pile shaft, preserving spatial correlations essential for geotechnical interpretation. Cross-validation results show cross validation errors below 3% across all loading stages, confirming strong prior predictive capability. In addition, the surrogate supports analytical global sensitivity analysis, offering a fast and reliable tool for identifying key parameter influences with minimal computational cost.

(ii) The proposed uncertainty treatment strategy enables robust and physically interpretable calibration. By jointly incorporating observation noise and surrogate error into an augmented covariance model, and addressing model discrepancy via spatial weighting, the framework avoids over-fitting and circumvents the need for arbitrary data or model

corrections. This leads to reliable posterior predictions that match field observations well, both in force responses across all loading stages and in displacement profiles near the mudline.

(iii) The integration of multiple field data sources enhances parameter calibration and improves overall model consistency. By assimilating both lateral deflections and pile applied forces through a custom-defined likelihood, the framework effectively fuses complementary information across construction stages. The sequential Bayesian updates progressively reduce uncertainty and reconcile the distinct sensitivities of individual data types, yielding stable and physically interpretable parameter estimates that remain consistent across observables.

5 Software and data availability

All codes related to sequential Bayesian inversion, uncertainties/error modelling and surrogate modelling are based on `Matlab` software. Markov Chain Monte Carlo is performed using an open-source toolbox UQLAB available at https://www.uqlab.com/. All scripts and datasets required to reproduce the figures are at https://github.com/ningxin999/Geotechnique.

Nomenclature

Greek symbols

α_T	Scaling factor for surrogate error
δ	Model error
ε_r	Observation noise
λ	Weight vector
ω	Polynomial coefficient
ψ	Polynomial basis
σ	Noise level
Σ_{ε_r}	Noise covariance
Σ_T	Surrogate error covariance
ξ	Model discrepancy

Latin symbols

D	Pile diameter
D_T	Total variance
$\mathcal{G}$	Observation group
G_0	Initial small-strain shear modulus
K	Number of realisations
K'	Number of test samples
K_0	Coefficient of earth pressure at rest
L	Pile length
M	Input space dimension
$\mathcal{M}$	Computational model
$\tilde{\mathcal{M}}$	Surrogate model
N	Output space dimension
N_{gr}	Number of observation groups

N'	Number of principal components
OCR	Overconsolidation ratio
S	Sobol indices
T	Surrogate error
v_G	Pile lateral displacement
x	Uncertain input soil parameters
y	FE output vector

Operators and sets

$\det(\cdot)$	Determinant
$\mathrm{diag}(\cdot)$	Diagonal matrix
$\emptyset$	Empty set
$\circ$	Hadamard product
$\mathbf{I}$	Identity matrix
$\mathcal{L}(\cdot)$	Likelihood function
$\pi(\cdot)$	Probability density function
$\mathbb{R}$	Set of real numbers
$(\cdot)^\top$	Transpose

References

Asem, P., & Gardoni, P. (2021). A generalized bayesian approach for prediction of strength and elastic properties of rock. *Engineering Geology*, *289*, 106187. https://doi.org/10.1016/j.enggeo.2021.106187

Baecher, G. B. (2017). Bayesian thinking in geotechnics. In *Geo-risk 2017* (pp. 1–18). https://doi.org/10.1061/9780784480694.001

Baecher, G. B. (2023). 2021 terzaghi lecture: Geotechnical systems, uncertainty, and risk. *Journal of Geotechnical and Geoenvironmental Engineering*, *149*(3), 03023001. https://doi.org/10.1061/JGGEFK.GTENG-1020

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Blatman, G., & Sudret, B. (2011). Principal component analysis and least angle regression in spectral stochastic finite element analysis. *Proc. 11th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP11), Zurich, Switzerland*. https://doi.org/10.1201/b11332-101

Brunton, S. L., & Kutz, J. N. (2022). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press. https://doi.org/10.1017/9781009089517

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual review of fluid mechanics*, *52*(1), 477–508. https://doi.org/10.1146/annurev-fluid-010719-060214

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse problems*, *30*(11), 114007. https://doi.org/10.1088/0266-5611/30/11/114007

Buckley, R., Chen, Y. M., Sheil, B., Suryasentana, S., Xu, D., Doherty, J., & Randolph, M. (2023). Bayesian optimization for cpt-based prediction of impact pile drivability. *Journal of Geotechnical and Geoenvironmental Engineering*, *149*(11), 04023100. https://doi.org/10.1061/JGGEFK.GTENG-11385

Byrne, B. W., McAdam, R. A., Burd, H. J., Beuckelaers, W. J. A. P., Gavin, K. G., Houlsby, G. T., Igoe, D. J. P., Jardine, R. J., Martin, C. M., Muir Wood, A., Potts, D. M., Skov Gretlund, J., Taborda, D. M. G., & Zdravkovic, L. (2020). Monotonic laterally loaded pile testing in a stiff glacial clay till at Cowden. *Géotechnique*, *70*(11), 970–985. https://doi.org/10.1680/jgeot.18.PISA.003

Christian, J. T. (2004). Geotechnical engineering reliability: How well do we know what we are doing? *Journal of geotechnical and geoenvironmental engineering*, *130*(10), 985–1003. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:10(985)

Conti, P., Guo, M., Manzoni, A., Frangi, A., Brunton, S. L., & Nathan Kutz, J. (2024). Multi-fidelity reduced-order surrogate modelling. *Proceedings of the Royal Society A*, *480*(2283), 20230655. https://doi.org/10.1098/rspa.2023.0655

Day, R., & Potts, D. (1994). Zero thickness interface elements—numerical stability and application. *International Journal for numerical and analytical methods in geomechanics*, *18*(10), 689–708. https://doi.org/10.1002/nag.1610181003

Garcia-Cabrejo, O., & Valocchi, A. (2014). Global sensitivity analysis for multivariate output using polynomial chaos expansion. *Reliability Engineering & System Safety*, *126*, 25–36. https://doi.org/10.1016/j.ress.2014.01.005

Geelen, R., Balzano, L., & Willcox, K. (2023). Learning latent representations in high-dimensional state spaces using polynomial manifold constructions. *2023 62nd IEEE Conference on Decision and Control (CDC)*, 4960–4965. https://doi.org/10.1109/CDC49753.2023.10384209

Hansen, T. M., Cordua, K. S., Jacobsen, B. H., & Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography. *Geophysics*, *79*(3), H1–H21. https://doi.org/10.1190/geo2013-0215.1

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hsein Juang, C., Luo, Z., Atamturktur, S., & Huang, H. (2013). Bayesian updating of soil parameters for braced excavations using field observations. *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(3), 395–406. https://doi.org/10.1061/(ASCE)GT.1943-5606.000078

Jin, Y., Biscontin, G., & Gardoni, P. (2021). Adaptive prediction of wall movement during excavation using bayesian inference. *Computers and Geotechnics*, *137*, 104249. https://doi.org/10.1016/j.compgeo.2021.104249

Kelly, R., & Huang, J. (2015). Bayesian updating for one-dimensional consolidation measurements. *Canadian Geotechnical Journal*, *52*(9), 1318–1330. https://doi.org/10.1139/cgj-2014-0338

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425–464. https://doi.org/10.1111/1467-9868.00294

Kramer, B., Peherstorfer, B., & Willcox, K. E. (2024). Learning nonlinear reduced models from data with operator inference. *Annual Review of Fluid Mechanics*, *56*(1), 521–548. https://doi.org/10.1146/annurev-fluid-121021-025220

Ling, Y., Mullins, J., & Mahadevan, S. (2014). Selection of model discrepancy priors in bayesian calibration. *Journal of Computational Physics*, *276*, 665–680. https://doi.org/10.1016/j.jcp.2014.08.005

Lo, M. K., & Leung, Y. (2019). Bayesian updating of subsurface spatial variability for improved prediction of braced excavation response. *Canadian Geotechnical Journal*, *56*(8), 1169–1183. https://doi.org/10.1139/cgj-2018-0409

McQuarrie, S. A., Huang, C., & Willcox, K. E. (2021). Data-driven reduced-order models via regularised operator inference for a single-injector combustion process. *Journal of the Royal Society of New Zealand*, *51*(2), 194–211. https://doi.org/10.1080/03036758.2020.1863237

Meles, G. A., Linde, N., & Marelli, S. (2022). Bayesian tomography with prior-knowledge-based parametrization and surrogate modelling. *Geophysical Journal International*, *231*(1), 673–691. https://doi.org/10.1093/gji/ggac214

Nagel, J. B., Rieckermann, J., & Sudret, B. (2020). Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation. *Reliability Engineering & System Safety*, *195*, 106737. https://doi.org/10.1016/j.ress.2019.106737

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V., & Alvin, K. F. (2002). Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, *75*(3), 333–357. https://doi.org/10.1016/S0951-8320(01)00120-X

Peherstorfer, B., & Willcox, K. (2016). Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, *306*, 196–215. https://doi.org/10.1016/j.cma.2016.03.025

Potts, D. M., & Zdravkovic, L. (1999). *Finite element analysis in geotechnical engineering: Theory*. Thomas Telford London. https://doi.org/10.1680/feaiget.27534

Qi, X.-H., & Zhou, W.-H. (2017). An efficient probabilistic back-analysis method for braced excavations using wall deflection data at multiple points. *Computers and Geotechnics*, *85*, 186–198. https://doi.org/10.1016/j.compgeo.2016.12.032

Rahimi, M., Shafieezadeh, A., Wood, D., Kubatko, E. J., & Dormady, N. C. (2019). Bayesian calibration of multi-response systems via multivariate kriging: Methodology and geological and geotechnical case studies. *Engineering geology*, *260*, 105248. https://doi.org/10.1016/j.enggeo.2019.105248

Schroeder, F., Day, R., Potts, D., & Addenbrooke, T. (2007). Quadrilateral isoparametric shear deformable shell element for use in soil-structure interaction problems. *International Journal of Geomechanics*, *7*(1), 44–52. https://doi.org/10.1061/(ASCE)1532-3641(2007)7:1(44)

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, *93*(7), 964–979. https://doi.org/10.1016/j.ress.2007.04.002

Tang, C., Cao, Z.-J., Hong, Y., & Li, W. (2022). State space model of undrained triaxial test data for bayesian identification of constitutive model parameters. *Géotechnique*, *74*(7), 697–711. https://doi.org/10.1680/jgeot.22.00144

Tian, H., Wang, Y., & Phoon, K.-K. (2024). Real-time fusion of multi-source monitoring data with geotechnical numerical model results using data-driven and physics-informed sparse dictionary learning. *Canadian Geotechnical Journal*, (ja). https://doi.org/10.1139/cgj-2023-0457

Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A., & Sudret, B. (2020). Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering structures*, *205*, 110063. https://doi.org/10.1016/j.engstruct.2019.110063

Wang, Y., Cao, Z., & Li, D. (2016). Bayesian perspective on geotechnical variability and site characterization. *Engineering Geology*, *203*, 117–125. https://doi.org/10.1016/j.enggeo.2015.08.017

Westermann, P., & Evins, R. (2019). Surrogate modelling for sustainable building design–a review. *Energy and buildings*, *198*, 170–186. https://doi.org/10.1016/j.enbuild.2019.05.057

Yang, N., Le, T., Zdravkovic, L., Potts, D., & Altmeyer, R. (2026). Sequential calibration of soil parameters using a two-step surrogate for high-dimensional geotechnical outputs. *Computers and Geotechnics*, *190*, 107688. https://doi.org/10.1016/j.compgeo.2025.107688

Yuen, K.-V., Ching, J., & Phoon, K.-K. (2021). Bayesian learning methods for geotechnical data. https://doi.org/10.1061/AJRUA6.0001102

Zdravkovic, L., Jardine, R. J., Taborda, D. M. G., Abadias, D., Burd, H. J., Byrne, B. W., Gavin, K. G., Houlsby, G. T., Igoe, D. J. P., Liu, T., Martin, C. M., McAdam, R. A., Muir Wood, A., Potts, D. M., Skov Gretlund, J., & Ushev, E. (2020a). Ground characterisation for PISA pile testing and analysis. *Géotechnique*, *70*(11), 945–960. https://doi.org/10.1680/jgeot.18.PISA.001

Zdravkovic, L., Taborda, D. M. G., Potts, D. M., Abadias, D., Burd, H. J., Byrne, B. W., Gavin, K. G., Houlsby, G. T., Jardine, R. J., Martin, C. M., McAdam, R. A., & Ushev, E. (2020b). Finite-element modelling of laterally loaded piles in a stiff glacial clay till at Cowden. *Géotechnique*, *70*(11), 999–1013. https://doi.org/10.1680/jgeot.18.PISA.005

Zheng, H., Mooney, M., & Gutierrez, M. (2023a). Surrogate model for 3d ground and structural deformations in tunneling by the sequential excavation method. *Computers and Geotechnics*, *154*, 105142. https://doi.org/10.1016/j.compgeo.2022.105142

Zheng, H., Mooney, M., & Gutierrez, M. (2023b). Updating model parameters and predictions in sem tunnelling using a surrogate-based bayesian approach. *Géotechnique*, *74*(13), 1855–1867. https://doi.org/10.1680/jgeot.22.00299

Zhou, W.-H., Yin, Z.-Y., & Yuen, K.-V. (2021). *Practice of bayesian probability theory in geotechnical engineering*. Springer. https://doi.org/doi.org/10.1007/978-981-15-9105-1