

Bayesian Calibration of Multi-Response Systems via Multivariate Kriging: Methodology and Geological and Geotechnical Case Studies

Mehrzaad Rahimi^a, Abdollah Shafieezadeh^{a,*}, Dylan Wood^a, Ethan Kubatko^a, and Noah C. Dormady^b

^aDep. of Civil, Env., and Geodetic Engineering, The Ohio State University, Columbus, OH, USA

^bJohn Glenn School of Public Affairs, The Ohio State University, Columbus, OH, USA

*Corresponding Author. Email address: shafieezadeh.1@osu.edu

Abstract

Monitoring systems often measure different types of responses of the built or natural environment at different points in space and time. Bayesian updating offers a robust method to use this information to update the distribution of involved uncertain system variables and enhance the predictive ability of the models in light of the new evidence. However, this process can be highly demanding and prohibitive in cases of sophisticated computational models. Here, we propose a highly efficient Bayesian updating framework that is integrated with multivariate Kriging surrogate modeling to quantify heteroscedastic uncertainties in the entire space of uncertain system variables and capture spatial and temporal dependencies among the responses using non-separable covariance structures. The advantages of the proposed framework are demonstrated on three geological and geotechnical examples, since responses in these systems are frequently multivariate in nature and geological properties are often highly uncertain. Results indicate that the developed framework is able to accurately and efficiently update uncertainties of system variables compared to existing Bayesian updating methods that are based on surrogate models. Furthermore, results show that considering the spatiotemporal dependencies between the responses in the framework can produce more accurate predictions of the future responses.

Keywords: Bayesian updating; probabilistic calibration; multivariate Kriging; spatiotemporal correlations; uncertainty quantification

1. Introduction

Accurate characterization of the performance and in particular the primary failure mechanisms of systems in the built or natural environments, *e.g.*, flood protection systems, transportation infrastructure, water and wastewater treatment facilities, and the power grid, is one of the main requirements to building an effective risk assessment framework. This objective may be achieved through three main components: computational models, uncertainty characterization, and reliability analysis. While all three components have significant impacts on the accuracy of the risk assessment framework, the second component, that is the probabilistic characteristics of involved system variables, can be particularly important for assessing geological and geotechnical systems, due to intricacies inherent to the geologic features (Christian, 2004). In this sense, many authors, using either deterministic methods (Adhikari et al., 2014; Sun et al., 2018) or probabilistic techniques (Zhang et al., 2010; Wang et al., 2013; Feng and Jimenez, 2015; Ering and Babu, 2016; Jiang et al., 2018; Yang et al., 2018; Sun et al., 2019; Wang et al., 2019) have addressed the analysis of uncertainty reduction in geo-mechanical properties through calibration. Whereas deterministic methods provide no information about the degree of uncertainty, probabilistic techniques of calibration provide the most likely updated distribution of the involved variables (Sun et al., 2018).

Among probabilistic techniques for calibration, Bayesian updating is shown to be a robust method for determining the probability density functions of system variables (Gelman et al., 2013; Zhou et al., 2017). Based on Bayes' theorem, the prior distribution of uncertain system variables to be updated and the knowledge learned from the observed data or measurements from full-scale experiments can be combined in a mathematical expression. This expression, which consists of prior distributions (from literature or engineering judgement) and a likelihood function (the chance to observe the response given system variables), very often cannot be solved analytically in the

case of multi-dimensional systems (*i.e.*, systems with multiple involved random variables). For solving the Bayesian formulation, Markov Chain Monte Carlo (MCMC) method is a sampling tool that has become increasingly practical (Gilks et al., 1995). As stated, the idea of MCMC is based on drawing samples from a proposal distribution in a chain, which requires thousands of evaluations of the likelihood function, which can be computationally expensive in the case of complex, computationally demanding numerical simulations, which is often the case in geological and geotechnical engineering. Moreover, utilizing various types of responses at different locations of a system during different time instances requires modification in the likelihood function that is embedded in the mathematical expression of Bayes' theorem (Zhang et al., 2018). To address the latter challenge, Juang et al. (2012) integrated two types of responses of a braced excavation, including maximum wall deflection and maximum ground settlement, in the likelihood function to simultaneously utilize them in Bayesian updating. Kelly and Huang (2015) also considered consolidation of a clay layer and then expressed the likelihood function for two types of responses, including settlement and excess pore water pressure, during various time instances using analytical formulations. Li et al. (2016c) combined incremental time-series monitoring data from multi-stage excavation of a high rock slope that was evaluated with a semi-empirical model into the likelihood function. These studies pointed out that incorporating more measured responses of different types and time instances yields higher accuracy in the posterior distribution of involved random variables. The predictive models in these studies were analytical formulations and semi-empirical models that are easy to evaluate; therefore, allowing thousands of simulations needed for Bayesian updating. However, the application of the methods in these studies to problems complex geological systems involving computationally demanding models may not be feasible.

Utilizing surrogate models, which are often statistically derived mathematical approximations of the ‘true’ models, can provide a general solution if integrated appropriately in the likelihood function. Several studies explored the integration of Response Surface Methodology (RSM), which is a polynomial regression model, for expressing the likelihood function in the probabilistic back-analysis process (Zhang et al., 2010; Miro et al., 2015; Li et al., 2016b; Qi and Zhou, 2017). Zhang et al. (2010) employed RSM as a surrogate model to approximate the safety factor of a slope calculated through limit equilibrium method. Then, they incorporated the surrogate model into the likelihood function and tuned the shear strength properties of the soil layers. They considered the potential error introduced by substituting the ‘true’ model with RSM as a random variable, and assumed it to follow a normal distribution with a constant standard deviation. This, however, is not a true representation of the uncertainty. In fact, the variance around the regression line is most often not constant for the entire space of system variables. Li et al. (2016b) utilized independent second-order polynomial response surfaces in lieu of the numerical model of a rock slope to approximate four responses (top vertical displacement, horizontal displacement of toe, depth of ground water table, and axial force of the rock bolt) and integrated the independent surrogate models into the likelihood function to update five involved system variables. In their approach, in addition to the error associated with the homoscedastic assumption for individual surrogate models, probabilistic associations among responses were entirely neglected. These drawbacks also exist in the vast majority of surrogate models other than RSM. Li et al. (2016a) adopted a multi-output support vector regression to approximate the displacements in a rock slope and integrated the surrogate model in the likelihood function to update distributions of geo-mechanical properties. Although the dependencies between the displacements at different locations were considered, similar to RSM, this method does not possess

the ability to accurately represent the uncertainty in the predictions. Sun et al., (2019) adopted independent Back Propagation (BP) neural networks as a surrogate model to determine displacements and anchorage forces of a rock slope during progressive excavation. However, the deviation of the surrogate model from measurements was assumed to stem from only the measurement error of the monitoring instruments, and the error of the surrogate models with respect to the computational model was entirely neglected. Despite improvements offered by previous studies, the current Bayesian updating frameworks that are integrated with surrogate models do not properly quantify uncertainties associated with the surrogate models nor the spatial and temporal dependencies between the responses of the systems of interest in the likelihood function. The probabilistic associations between responses of different or the same types are often directly related to the proximity of the data in space and time specifically in geological and geotechnical systems.

Based on the above discussion, it is of considerable value to develop the most general formulation of the likelihood function for multi-response physical phenomena using multivariate surrogate models that can accurately quantify the uncertainties and capture the associations among responses. In this study, we propose integrating into Bayesian updating a multivariate Kriging surrogate model that can capture heteroscedastic uncertainties over the entire space of system variables and moreover consider the probabilistic associations between responses. The multivariate Kriging model is constructed based on a multi-objective optimization method developed by Svenson (2011). The proposed representation of the likelihood function is integrated with Bayes' theorem and solved using the MCMC algorithm. The merits of this approach are evaluated for three geological and geotechnical systems with multi-dimensional responses and soil properties that are highly uncertain. The proposed method is elaborated in Section 2, which

contains two subsections detailing the formulation of the likelihood function with multivariate Kriging and MCMC algorithm. Section 3 begins with a summary of the goals pursued in the application examples, and then provides the examples with a detailed analysis of results. Finally, concluding remarks are presented in the Conclusion section.

2. Methodology

In this section, we present the details of the proposed Bayesian updating and an extended formulation for probabilistic calibration for multiple responses, which is built by incorporating multivariate Gaussian process regression (Kriging) within the expression for the likelihood function. Use of the Kriging meta-modeling technique provides a means for considering non-homoscedastic model uncertainty over the input variable space. Furthermore, since responses in the likelihood function are generally from different types of physical quantities, in this method, the covariance matrix of the likelihood function considers both the associations over the input variable space and the associations among the responses by a non-separable covariance structure. Subsequently, the issues that arise during implementation of the Bayesian updating formulation for multi-dimensional systems and the sampling method (Markov Chain Monte Carlo, MCMC) that is used to overcome these issues are discussed.

2.1 Bayesian updating for multiple responses of different types of physical quantities

The Bayesian method for probabilistic calibration statistically integrates information gathered, about a phenomenon or a system, whose behavior is captured by a predictive model, to arrive at the most likely representation of involved uncertain variables. Indeed, prior probabilistic information of uncertain system variables, which is assumed from available measurements and/or expert belief, may be updated by the Bayesian method to posterior probabilistic information, based on new information. Such information may be in the form of field observations (qualitative

information) and/or measurements (quantitative information). The probability of observing new information can be quantified using a likelihood function, and Bayesian analysis then allows updating the probabilistic belief about the variables given the likelihood of the new information. Let us consider θ as the single event of interest, then based on Bayes' theorem, the probability of θ conditioned on observing data M can be defined as follows,

$$P(\theta|M) = \frac{P(\theta \cap M)}{P(M)} = \frac{P(M|\theta)P(\theta)}{P(M)}, \quad (1)$$

where $P(\theta|M)$ is the posterior probability of θ , $P(\theta)$ is the prior probability of θ , $P(M|\theta)$ is the likelihood function, and $P(M)$ is the probability of observing M . In fact, the observing data are characterized by the likelihood function, and Bayes' theorem guides us on how to update the prior probability information to the posterior probability information using the likelihood function. The application of Bayes' theorem for updating of a single event θ in Equation (1) can be extended to the calibration of a d -dimensional set of involved variables, $\theta = [\theta_1, \dots, \theta_t, \dots, \theta_d]$ as follows,

$$f''_{\theta}(\theta) = \frac{L(\theta)f'_{\theta}(\theta)}{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\theta)f'_{\theta}(\theta) d\theta_1 \dots d\theta_d} = \frac{L(\theta)f'_{\theta}(\theta)}{\int_{\theta} L(\theta)f'_{\theta}(\theta) d\theta} \quad (2)$$

where $f'_{\theta}(\theta)$ and $f''_{\theta}(\theta)$ are the prior and posterior Probability Density Functions (PDFs) of θ , respectively, and $L(\theta)$ is the likelihood function that describes the occurrence probability of measurements and/or observations. In Equation (2), $1/\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\theta)f'_{\theta}(\theta) d\theta_1 \dots d\theta_d = c$ is a normalization constant that assures the posterior distribution integrates to unity. Thereby, Equation (2) is simplified as follows,

$$f''_{\theta}(\theta) = cL(\theta)f'_{\theta}(\theta). \quad (3)$$

In order to introduce the likelihood function in the case of multiple responses, it is helpful to specify the following definitions: we will use the notation y_{jk} to indicate the particular value of the k^{th} response that is measured at the j^{th} time instance. Consequently, the p measured responses

159 $(k \in \{1, \dots, p\})$ from t_1 to t_n ($j \in \{1, \dots, n\}$) can be written as a stacked vector, called $\mathbf{Y}$, of $m =$
 160 $n \cdot p$ rows as follows,

$$\mathbf{Y} = [\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_j^T, \dots, \mathbf{Y}_n^T]^T_{m \times 1}, \quad (4)$$

161 where $\mathbf{Y}_j^T = [y_{j1}, y_{j2}, \dots, y_{jk}, \dots, y_{jp}]_{1 \times p}$ is the p -dimensional vector of the measured responses at the
 162 j^{th} time instance. Let $h_{jk}(\boldsymbol{\theta})$ denote the simulation response corresponding to y_{jk} . Similarly, the
 163 corresponding stacked vector of the simulation responses, $\mathbf{H}(\boldsymbol{\theta})$, is written as follows,

$$\mathbf{H}(\boldsymbol{\theta}) = [\mathbf{H}_1^T(\boldsymbol{\theta}), \mathbf{H}_2^T(\boldsymbol{\theta}), \dots, \mathbf{H}_j^T(\boldsymbol{\theta}), \dots, \mathbf{H}_n^T(\boldsymbol{\theta})]^T_{m \times 1}. \quad (5)$$

164 In Equation (5), $\mathbf{H}_j^T(\boldsymbol{\theta}) = [h_{j1}(\boldsymbol{\theta}), h_{j2}(\boldsymbol{\theta}), \dots, h_{jk}(\boldsymbol{\theta}), \dots, h_{jp}(\boldsymbol{\theta})]_{1 \times p}$ is the p -dimensional vector
 165 of the simulation responses at the j^{th} time instance. Let the stacked vector $\boldsymbol{\epsilon}(\boldsymbol{\theta})$ denote the
 166 deviations of the simulation responses from the measured responses, which are either due to
 167 measurement errors, uncertainties in input system variables, the degree of fidelity of the
 168 computational models, or some combination of these sources. In the case of additive errors, the
 169 response vector is represented as,

$$\mathbf{Y} = \mathbf{H}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}(\boldsymbol{\theta}). \quad (4)$$

170 Therefore, the likelihood function describing the multivariate PDF of the deviation is as follows,

$$L_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}) = f_{\boldsymbol{\epsilon}}(\mathbf{Y} - \mathbf{H}(\boldsymbol{\theta})). \quad (5)$$

171 By assuming the likelihood function to be a multivariate normal distribution with the $m \times m$
 172 covariance matrix of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ that denotes the correlations between the entries of the deviation vector,
 173 $\boldsymbol{\epsilon}(\boldsymbol{\theta})$, $L_{\boldsymbol{\epsilon}}(\boldsymbol{\theta})$ takes the following form,

$$L_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}} (|\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}|)^{\frac{1}{2}}} \times \exp \left[-\frac{1}{2} [\mathbf{Y} - \mathbf{H}(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} [\mathbf{Y} - \mathbf{H}(\boldsymbol{\theta})] \right], \quad (6)$$

where $|\Sigma_\varepsilon|$ is the determinant of the deviation covariance matrix. Each entry of the deviation covariance matrix denotes the dependency between the errors associated with the corresponding responses. This error is related to the accuracy of the measurements (e.g. signal to noise ratio of data measured by sensors) and fidelity of the computational models. Often measurement systems are calibrated to work properly for a specific range of responses and therefore their accuracy may not be uniform across a wide range of responses (Khosravi et al., 2017). Additionally, it is often the case that the confidence in the responses of the computational models may not remain the same over the output space, and therefore the error of the models may not be uniform when expressed as a function of model variables. For instance, when the input system variables impose a non-linear system behavior, numerical simulations may not provide results that are as accurate as for when the system behaves linearly (Matthies et al., 1997). These variations in the uncertainties that are associated with measurement systems and computational models can be incorporated in the deviation covariance matrix, if they can be quantified accurately.

Simulated responses, $\mathbf{H}(\boldsymbol{\theta})$, can be approximated statistically using a stochastic surrogate model, $\hat{\mathbf{K}}(\boldsymbol{\theta})$, which is especially useful for the case where $\mathbf{H}(\boldsymbol{\theta})$ is obtained by a computationally expensive numerical model. Similar to $\mathbf{Y}$ and $\mathbf{H}(\boldsymbol{\theta})$, we define $\hat{\mathbf{K}}(\boldsymbol{\theta})$ as a stacked vector that includes m entries, *i.e.*,

$$\hat{\mathbf{K}}(\boldsymbol{\theta}) = [\hat{\mathbf{K}}_1^T(\boldsymbol{\theta}), \hat{\mathbf{K}}_2^T(\boldsymbol{\theta}), \dots, \hat{\mathbf{K}}_j^T(\boldsymbol{\theta}), \dots, \hat{\mathbf{K}}_n^T(\boldsymbol{\theta})]_{m \times 1}^T, \quad (9)$$

where, $\hat{\mathbf{K}}_j^T(\boldsymbol{\theta}) = [\hat{k}_{j1}(\boldsymbol{\theta}), \hat{k}_{j2}(\boldsymbol{\theta}), \dots, \hat{k}_{jk}(\boldsymbol{\theta}), \dots, \hat{k}_{jp}(\boldsymbol{\theta})]_{1 \times p}$ is the p -dimensional vector of the surrogate model responses at the j^{th} time instance. Then the response vector, Equation (6), is extended as:

$$\mathbf{Y} = \mathbf{H}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}(\boldsymbol{\theta}) = \hat{\mathbf{K}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}(\boldsymbol{\theta}) = \mathbb{E}[\hat{\mathbf{K}}(\boldsymbol{\theta})] + \boldsymbol{\delta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}(\boldsymbol{\theta}), \quad (10)$$

where $E[\hat{\mathbf{K}}(\boldsymbol{\theta})]$ corresponds to the mean values of the surrogate model predictions for each response, and the prediction errors, $\boldsymbol{\delta}(\boldsymbol{\theta})$, account for the discrepancies between the surrogate model predictions and those by the ‘true’ model. In addition to the mean values of the surrogate model predictions, the prediction error of the surrogate model, as shown in Equation (10), directly influences the response vector that plays an essential role in the likelihood function and hence the posterior distributions. In the vast majority of surrogate models, such as RSM, the prediction error is assumed to be homoscedastic, which implies a uniform variance over the entire space of input variables, $\boldsymbol{\theta}$. This assumption is not close to reality, as predictions for samples that are close to the design samples of the surrogate model will have higher confidences compared with those that are far away from the design samples (Rahimi et al., 2019). In the cases where multiple responses of different types of physical quantities are involved, some surrogate models reduce multi-dimensional responses of systems to a single response, either by applying weighting functions or by considering only the most significant response. Alternatively, in these cases, some approaches construct a univariate surrogate model for each response of interest, and therefore neglect correlations among responses. The most appropriate way to build a surrogate model is by a multivariate model that considers heteroscedastic uncertainty over the entire space of input variables as well as the probabilistic associations between responses of different types of quantities.

Here, we propose to consider such associations in the formulation of the likelihood function through multivariate Gaussian process regression, known as multivariate Kriging. This model has a constant mean vector of $\boldsymbol{\beta} = E[\hat{\mathbf{K}}(\boldsymbol{\theta})]$ and a $m \times m$ covariance matrix of $\boldsymbol{\Sigma}_{\delta}$ that denotes the dependencies between the entries of $\boldsymbol{\delta}(\boldsymbol{\theta})$. Since $\boldsymbol{\Sigma}_{\delta}$ captures two types of correlation: correlations among responses of one type over the input variable space and correlations between different

217 responses, determining its form can be challenging. In some cases where only one type of physical
 218 quantity is involved in the responses (*e.g.*, displacement of a cantilever beam at different locations
 219 of the beam), it may be appropriate to treat the two kinds of correlation as separable and to then
 220 determine the covariance structure. However, when multiple responses arise from different types
 221 of quantities, a separable covariance structure is no longer appropriate, since there is no obvious
 222 index between the responses (Fricker et al., 2010). Fricker et al. (2010) showed that separable
 223 covariance structures have a Markovian type property. For instance, if we have two types of
 224 responses (A and B) and desire estimating response A at location X , and we have already observed
 225 response A at location Y , then observation of response B at location Y gives us no further
 226 information. Therefore, to arrive at a general formulation for the likelihood function, here we
 227 define Σ_δ with a non-separable covariance structure by using a method known as Non-Linear
 228 Model of Coregionalization (NLMC) (Fricker et al., 2010). Using NLMC, $\hat{\mathbf{K}}(\boldsymbol{\theta})$ is expressed as a
 229 process of the form:

$$\hat{\mathbf{K}}(\boldsymbol{\theta}) = \mathbf{E}[\hat{\mathbf{K}}(\boldsymbol{\theta})] + \boldsymbol{\delta}(\boldsymbol{\theta}) = \boldsymbol{\beta} + \mathbf{A}\mathbf{Z}(\boldsymbol{\theta}), \quad (11)$$

230 where $\boldsymbol{\beta} = [\beta_1, \dots, \beta_i, \dots, \beta_m]^T$ contains the constant mean values for each response of the
 231 numerical simulation (*i.e.*, $\mathbf{E}[\hat{\mathbf{K}}(\boldsymbol{\theta})] = \boldsymbol{\beta}$); $\mathbf{Z}(\boldsymbol{\theta}) = [Z_1(\boldsymbol{\theta}), \dots, Z_i(\boldsymbol{\theta}), \dots, Z_m(\boldsymbol{\theta})]^T$ is an m -
 232 dimensional vector of mutually independent stationary Gaussian processes with zero mean, unit
 233 variance, and a Gaussian correlation function; and $\mathbf{A}$ is a $m \times m$ symmetric, positive-definite
 234 matrix that is the square root of Σ_δ computed in the case of two random system variable vectors
 235 of the same values. In more details, the Gaussian correlation function for $Z_i(\boldsymbol{\theta})$ is as follows:

$$R_i(\boldsymbol{\theta}', \boldsymbol{\theta}'', \boldsymbol{\eta}_i) = \prod_{t=1}^d \exp(-\eta_{i,t}(\theta'_t - \theta''_t)^2), \quad (12)$$

236 where, $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ are two arbitrary d -dimensional vectors of random system variables, and where
 237 $\boldsymbol{\eta}_i$ is a d -dimensional vector of correlation coefficients for each response. It is straightforward to
 238 see that Equation (11) together with Equation (12) implies:

$$239 \quad \mathbf{COV}(\hat{\mathbf{K}}(\boldsymbol{\theta}'), \hat{\mathbf{K}}(\boldsymbol{\theta}'')) = \mathbf{A} \text{diag}(R_1(\boldsymbol{\theta}', \boldsymbol{\theta}'', \boldsymbol{\eta}_1), \dots, R_i(\boldsymbol{\theta}', \boldsymbol{\theta}'', \boldsymbol{\eta}_i), \dots, R_m(\boldsymbol{\theta}', \boldsymbol{\theta}'', \boldsymbol{\eta}_m)) \mathbf{A}^T, \quad (13)$$

240 where $\mathbf{COV}(\hat{\mathbf{K}}(\boldsymbol{\theta}'), \hat{\mathbf{K}}(\boldsymbol{\theta}''))$ is the $m \times m$ covariance matrix between the multiple responses, and
 241 $\mathbf{A}$ is defined through the eigen-decomposition of $\mathbf{COV}(\hat{\mathbf{K}}(\boldsymbol{\theta}'), \hat{\mathbf{K}}(\boldsymbol{\theta}''))$ when $\boldsymbol{\theta}' = \boldsymbol{\theta}''$.

242 There are $d \cdot m$ and $m \cdot (m + 1)/2$ covariance coefficients in $\boldsymbol{\eta}$ and $\mathbf{A}$, respectively. Here,
 243 we implement the multi-objective optimization method developed by Svenson (2011) to obtain
 244 Restricted Maximum Likelihood (REML) estimates of the covariance coefficients in $\mathbf{A}$ and $\boldsymbol{\eta}$. The
 245 adopted optimization process not only fits the model at the initial design of experiment (*i.e.*,
 246 estimation of $\mathbf{A}$ and $\boldsymbol{\eta}$ in the covariance matrix), but also obtains the predictions, $E[\hat{\mathbf{K}}(\boldsymbol{\theta})]$, and
 247 prediction errors, $\boldsymbol{\delta}(\boldsymbol{\theta}) = \mathbf{AZ}(\boldsymbol{\theta})$, at any desired sample. This probabilistic information is used to
 248 modify the likelihood function based on the multivariate surrogate model as follows,

$$L_{\varepsilon\delta}(\boldsymbol{\theta}) = f_{\varepsilon\delta}(\mathbf{Y} - E[\hat{\mathbf{K}}(\boldsymbol{\theta})]). \quad (14)$$

249 Equation (14) delineates that the covariance matrix in the likelihood function is now associated
 250 with both $\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ and $\boldsymbol{\delta}(\boldsymbol{\theta})$, and that likelihood function will have the following form,

$$L_{\varepsilon\delta}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}} (|\boldsymbol{\Sigma}_{\varepsilon\delta}|)^{\frac{1}{2}}} \times \exp \left[-\frac{1}{2} [\mathbf{Y} - E[\hat{\mathbf{K}}(\boldsymbol{\theta})]]^T \boldsymbol{\Sigma}_{\varepsilon\delta}^{-1} [\mathbf{Y} - E[\hat{\mathbf{K}}(\boldsymbol{\theta})]] \right], \quad (15)$$

251 The new likelihood function with the covariance matrix $\boldsymbol{\Sigma}_{\varepsilon\delta} = \boldsymbol{\Sigma}_{\varepsilon} + \boldsymbol{\Sigma}_{\delta}$ that considers all the
 252 uncertainties regarding errors associated with the measurements systems, computational models,
 253 and surrogate models, presents the most general formulation of the likelihood function, where

probabilistic associations among deviations of different types of responses in space and time are accounted for.

In order to show the general form of the covariance array, Σ_δ , as an example, we assume $p = 2$ (two responses) and $n = 3$ (three time instances); therefore, $m = 3 \cdot 2 = 6$, which means that the covariance matrix is 6×6 . Since $n = 3$ we have $3^2 = 9$ block matrices in the entire covariance matrix, and since $p = 2$, those 9 matrices are 2×2 . Figure 1 provides an illustration of the covariance matrix for three different cases.

Schematic view

(a)

(b)

(c)

y_{jk} : k^{th} response at j^{th} time

$$\begin{aligned}
 & \begin{array}{c} Y_{11} \quad Y_{21} \quad Y_{31} \\ Y_{12} \quad Y_{22} \quad Y_{32} \end{array} \quad \begin{array}{c} Y_{11} \quad Y_{12} \quad Y_{21} \quad Y_{22} \quad Y_{31} \quad Y_{32} \\ \begin{bmatrix} \sigma_{11}(\theta)\sigma_{11}(\theta) & \sigma_{11}(\theta)\sigma_{12}(\theta) & \sigma_{11}(\theta)\sigma_{21}(\theta) & \sigma_{11}(\theta)\sigma_{22}(\theta) & \sigma_{11}(\theta)\sigma_{31}(\theta) & \sigma_{11}(\theta)\sigma_{32}(\theta) \\ \sigma_{12}(\theta)\sigma_{11}(\theta) & \sigma_{12}(\theta)\sigma_{12}(\theta) & \sigma_{12}(\theta)\sigma_{21}(\theta) & \sigma_{12}(\theta)\sigma_{22}(\theta) & \sigma_{12}(\theta)\sigma_{31}(\theta) & \sigma_{12}(\theta)\sigma_{32}(\theta) \\ \sigma_{21}(\theta)\sigma_{11}(\theta) & \sigma_{21}(\theta)\sigma_{12}(\theta) & \sigma_{21}(\theta)\sigma_{21}(\theta) & \sigma_{21}(\theta)\sigma_{22}(\theta) & \sigma_{21}(\theta)\sigma_{31}(\theta) & \sigma_{21}(\theta)\sigma_{32}(\theta) \\ \sigma_{22}(\theta)\sigma_{11}(\theta) & \sigma_{22}(\theta)\sigma_{12}(\theta) & \sigma_{22}(\theta)\sigma_{21}(\theta) & \sigma_{22}(\theta)\sigma_{22}(\theta) & \sigma_{22}(\theta)\sigma_{31}(\theta) & \sigma_{22}(\theta)\sigma_{32}(\theta) \\ \sigma_{31}(\theta)\sigma_{11}(\theta) & \sigma_{31}(\theta)\sigma_{12}(\theta) & \sigma_{31}(\theta)\sigma_{21}(\theta) & \sigma_{31}(\theta)\sigma_{22}(\theta) & \sigma_{31}(\theta)\sigma_{31}(\theta) & \sigma_{31}(\theta)\sigma_{32}(\theta) \\ \sigma_{32}(\theta)\sigma_{11}(\theta) & \sigma_{32}(\theta)\sigma_{12}(\theta) & \sigma_{32}(\theta)\sigma_{21}(\theta) & \sigma_{32}(\theta)\sigma_{22}(\theta) & \sigma_{32}(\theta)\sigma_{31}(\theta) & \sigma_{32}(\theta)\sigma_{32}(\theta) \end{bmatrix} \end{array} \\
 & \begin{array}{c} Y_{11} \quad Y_{21} \quad Y_{31} \\ Y_{12} \quad Y_{22} \quad Y_{32} \end{array} \quad \begin{array}{c} Y_{11} \quad Y_{12} \quad Y_{21} \quad Y_{22} \quad Y_{31} \quad Y_{32} \\ \begin{bmatrix} \sigma_{11}(\theta)\sigma_{11}(\theta) & \sigma_{11}(\theta)\sigma_{12}(\theta) & 0 & 0 & 0 & 0 \\ \sigma_{12}(\theta)\sigma_{11}(\theta) & \sigma_{12}(\theta)\sigma_{12}(\theta) & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{21}(\theta)\sigma_{21}(\theta) & \sigma_{21}(\theta)\sigma_{22}(\theta) & 0 & 0 \\ 0 & 0 & \sigma_{22}(\theta)\sigma_{21}(\theta) & \sigma_{22}(\theta)\sigma_{22}(\theta) & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{31}(\theta)\sigma_{31}(\theta) & \sigma_{31}(\theta)\sigma_{32}(\theta) \\ 0 & 0 & 0 & 0 & \sigma_{32}(\theta)\sigma_{31}(\theta) & \sigma_{32}(\theta)\sigma_{32}(\theta) \end{bmatrix} \end{array} \\
 & \begin{array}{c} Y_{11} \quad Y_{21} \quad Y_{31} \\ Y_{12} \quad Y_{22} \quad Y_{32} \end{array} \quad \begin{array}{c} Y_{11} \quad Y_{12} \quad Y_{21} \quad Y_{22} \quad Y_{31} \quad Y_{32} \\ \begin{bmatrix} \sigma_{11}(\theta)\sigma_{11}(\theta) & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{12}(\theta)\sigma_{12}(\theta) & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{21}(\theta)\sigma_{21}(\theta) & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{22}(\theta)\sigma_{22}(\theta) & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{31}(\theta)\sigma_{31}(\theta) & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{32}(\theta)\sigma_{32}(\theta) \end{bmatrix} \end{array}
 \end{aligned}$$

Figure 1. The covariance matrix of $\delta(\theta)$ (Σ_δ) for an example with $p = 2$ and $n = 3$; (a) fully dependent responses, (b) independent responses in time, and (c) fully independent responses.

In this figure, the grey boxes in the schematic view denote the dependencies between the responses. As illustrated in Figure 1a, the 6×6 covariance matrix has all non-zero entries in the case of considering all the dependencies between different types of responses during different time instances. In the case of considering the dependencies between the types of responses at each time instance, but independent of other time instances (Figure 1b), the off-diagonal 2×2 block matrices are zero. Finally, one can consider all of the responses to be independent of each other, which yields a strictly diagonal (i.e., not just block diagonal, as in Figure 1b) covariance matrix, as shown in Figure 1c.

2.2 Markov Chain Monte Carlo (MCMC)

Techniques such as conjugate prior, direct integration method, and Maximum Posterior Density (MPD) serve as analytical and approximation tools for exploring the posterior distribution through solving Equation (3); however, solving this equation can be highly demanding for the general likelihood function in Equation (15). Therefore, in this case one needs to employ a sampling method, such as MCMC, in order to solve the Bayes' equation. In this sub-section, the details of MCMC are provided. The concept of Monte Carlo integration, combined with that of a Markov Chain (a mathematical description for a stochastic sequential process), yields a strong sampling method called Markov Chain Monte Carlo (MCMC) that can be used to solve the Bayes' equation (Equation 3). MCMC aims to build a Markov Chain with a histogram similar to the target distribution, which in the case of this study is the posterior distribution of the involved system variables, $f''_{\theta}(\theta)$, in Equation (3). MCMC aims to converge to the target distribution irrespective of the initial values in the chain. To achieve this goal, there are several sequential sampling algorithms including Metropolis-Hastings (MH), Metropolis, and Gibbs sampler that vary in the

287 definition of the transition process in the chain (Gelman et al., 2013). Here, we use the Metropolis
 288 sampler, which produces the following sequential states,

$$\boldsymbol{\theta}^{(1)} \rightarrow \boldsymbol{\theta}^{(2)} \rightarrow \dots \rightarrow \boldsymbol{\theta}^{(r)} \rightarrow \dots \rightarrow \boldsymbol{\theta}^{(R)}, \quad (16)$$

289 where R is the length of the Markov Chain and $\boldsymbol{\theta}^{(r)}$ is the r^{th} state in the chain. The first state of
 290 the chain $\boldsymbol{\theta}^{(1)} = [\theta_1^{(1)}, \dots, \theta_i^{(1)}, \dots, \theta_d^{(1)}]$ (d is the dimension of the involved system variables), is
 291 randomly selected from the logical ranges of each θ_i . A candidate vector, $\boldsymbol{\theta}^{(*)}$, is produced
 292 subsequently from the conditional proposal distribution $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r-1)})$. In this study $q(\cdot)$ is a
 293 multivariate normal distribution with a mean of $\boldsymbol{\theta}^{(r-1)}$ and the covariance matrix of $\zeta\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, where ζ
 294 is a scaling factor and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the covariance matrix of the prior distribution. To decide whether or
 295 not to accept the candidate vector, $\boldsymbol{\theta}^{(*)}$, a uniform random number, u , in the interval of zero and
 296 one is generated and compared with the probability of accepting the proposal, α , which in the case
 297 of Metropolis sampler is defined as follows,

$$\alpha = \min\left(1, \frac{f''(\boldsymbol{\theta}^{(*)})}{f''(\boldsymbol{\theta}^{(r-1)})}\right) = \min\left(1, \frac{cL(\boldsymbol{\theta}^{(*)})f'(\boldsymbol{\theta}^{(*)})}{cL(\boldsymbol{\theta}^{(r-1)})f'(\boldsymbol{\theta}^{(r-1)})}\right). \quad (17)$$

298 If $u \leq \alpha$, the candidate vector is accepted and the next state is set as $\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(*)}$. Otherwise, the
 299 candidate is rejected and the next state is set as $\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(r-1)}$. This process continues for the entire
 300 length of the chain, R . Since the initial samples in the Markov Chain are unstable, they should be
 301 discarded from the chain. In this study, R is selected to be 10000, and the first 1000 samples, which
 302 are called ‘burnt’ samples, are discarded.

303 3. Application examples

304 To evaluate the proposed method, three different geotechnical examples with different dimensions
 305 in random system variables and various types of measured quantities are considered in this study.
 306 Although details of each example with the corresponding results are presented and discussed in

detail in the following sub-sections, Table 1 illustrates a summary of each example to clarify the goals we are perusing for each example.

Table 1: Summary of application examples

#	Type of example	Measured quantity	p	n	d	Goals
1	Homogenous soil slope	• Factor of safety	1	1	3	• Verification of the method for single response ($m = 1$) • Impact of homoscedasticity on the prediction error
2	Consolidation of a clay layer	• Settlement • Pore pressure	2	3	4	• Verification of the method for multiple responses ($m \geq 1$) • Impact of multi-type responses • Impact of surrogate type • Impact of training sample numbers
3	Water rise in a levee	• Wall displacement • Pore pressure 1 • Pore pressure 2	3	3,4	8	• Impact of high model complexity • Impact of high dimensions

3.1 Example 1: Homogenous soil slope

In this example, a three-layered homogenous soil slope was cut back to a steeper design, which resulted in the slope to fail. Figure 2 shows the steepened geometry and the soil layer profiles of the slope. Duncan (1999) reported the mean values of the shear strength properties of soil layers, which are summarized in Table 2. Using these mean values and the strength reduction method through FLAC3D (Itasca, 2017), we determine the safety factor of the slope as 1.41, which is consistent with the reported safety factor in (Duncan, 1999; Zhang et al., 2010); however, it contradicts the true observed failure. This discrepancy motivated Duncan (1999) and Zhang et al. (2010) to respectively perform deterministic and probabilistic calibrations in order to update the shear strength properties of the soil layers with the assumption that the slope failure is initiated by the change in the slope steepness, not changes in the soil properties. Zhang et al. (2010) considered the shear strength properties of the top soil layers (sandy clay and highly plastic clay) as uncertain variables and then based on literature and engineering judgement they assumed the prior probabilistic models that are elaborated in Table 2. They constructed a second order polynomial

function with seven design samples and then evaluated the accuracy of the surrogate model through comparing safety factor determined by the surrogate model and numerical simulations on sixteen more design samples. They subsequently performed MCMC with the likelihood function integrated with the constructed surrogate model. Although not discussed in their work, their treatment of the uncertainty in light of the general likelihood structure presented in this research is to mix the error associated with the numerical simulation, $\varepsilon(\theta)$, and the potential error introduced by substituting the true model with RSM, $\delta(\theta)$. The resulting error is considered to have a normal distribution with a constant non-zero mean and constant standard deviation (*i.e.*, $\delta(\theta) + \varepsilon(\theta) = \varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon)$). The main issue in that method is the assumption of homoscedasticity, which is not the true representation of the uncertainty.

In this study, similar to Zhang et al. (2010), we assume that the slope failure is originated by the change in the slope steepness. With this assumption, we seek to find the updated soil properties given that the slope has failed, *i.e.*, given that factor of safety is unity. Similar to Zhang et al. (2010) we consider friction angle and cohesion of sandy clay, and friction angle of highly plastic clay as random variables that we are interested in updating through the proposed method. Therefore, d equals to three and $\theta = [c_1, \varphi_1, \varphi_2]$, where $c_1, \varphi_1, \varphi_2$ are defined in Table 2. We also assume that these input system variables are statistically independent and follow lognormal distributions with the statistical properties shown in Table 2.

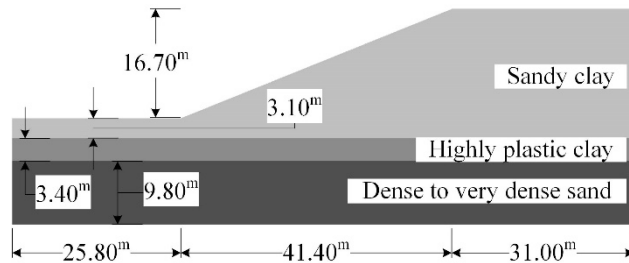


Figure 2. Cross section of soil slope in Example 1.

Table 2: Probabilistic models of slope properties in Example 1.

Soil Parameters	Distribution	Mean (μ)	C.O.V
Cohesion of sandy clay, c_1 (kPa)	Lognormal	14.4	0.20
Friction angle of sandy clay, φ_1 (°)	Lognormal	35.0	0.15
Friction angle of highly plastic clay, φ_2 (°)	Lognormal	25.0	0.15

Figures 3a, 3b, and 3c show the MCMC iteration samples with the posterior histograms for cohesion of sandy clay, friction angle of sandy clay, and friction angle of highly plastic clay. Among the crucial implementation issues of MCMC, the choices for the number of iterations and for the starting value are of high importance. The bias caused by the latter can be reduced by discarding a number of the initial samples, which are considered as belonging to the initial transient phase of the method (Brooks and Roberts, 1998).

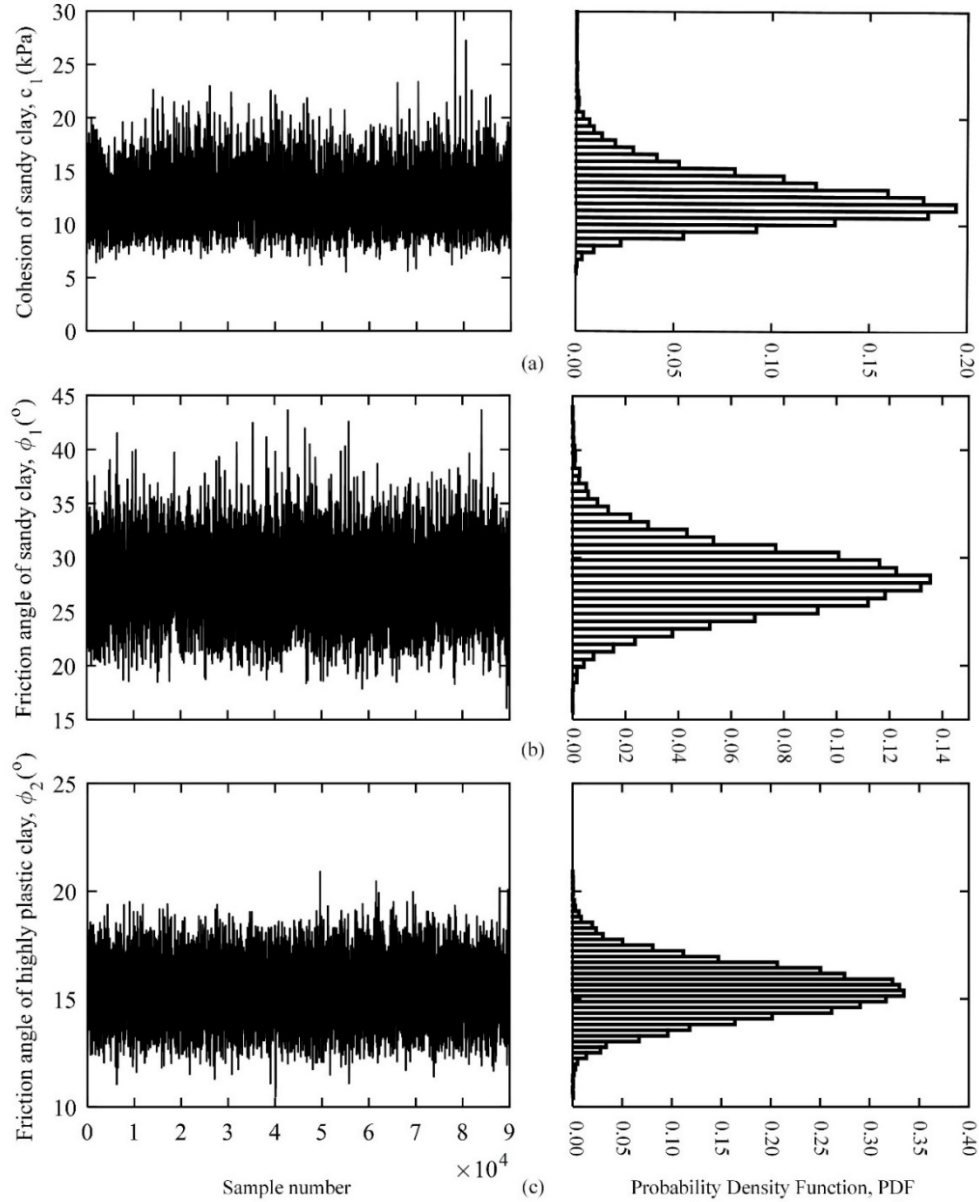


Figure 3. MCMC iterations and posterior histograms for (a) cohesion of sandy clay, c_1 ; (b) friction angle of sandy clay, ϕ_1 ; (c) friction angle of highly plastic clay, ϕ_2 .

In order to choose a sufficient number of iterations to achieve convergence, various methods exist, such as making use of autocorrelation plots as suggested by Gelfand and Smith (1990). Here, the number of samples, R , is chosen to be 10000 and the first 1000 samples are “burnt”. Autocorrelation tests show that the MCMC chain is stabilized. The scaling factor in the proposal distribution is $\zeta = 0.5$, which yields to the acceptance rate of 0.33, as suggested by Gelman et al.

(2013). As indicated by Figure 3, the posterior mean values of the soil strength parameters are decreased compared with the prior mean values (Table 2), which is logical, since the prior safety factor (1.41) was overestimated. Using the updated properties in the FLAC3D simulation yields a factor of safety that is very close to one, which is consistent with the field observation.

For the next phase of verification, the correlation between cohesion and friction angle of sandy clay, using the prior and posterior distributions are illustrated in Figure 4a and 4b, respectively. While Figure 4a shows no correlation between cohesion and friction angle, there is a negative correlation between them in Figure 4b, which is consistent with previous studies (Li et al., 2015; Wang and Akeju, 2016) that show the negative correlation between these two variables is a physical feature of soils.

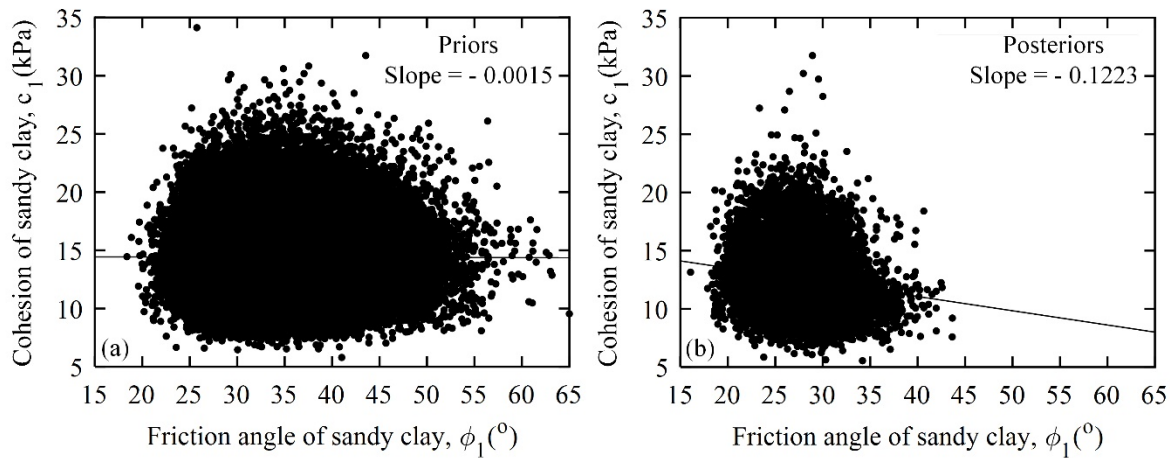


Figure 4. correlation between the friction angle and cohesion of sandy clay (a) priors, (b) posteriors.

The posterior histograms for all three random variables using the proposed method are compared with the results from Zhang et al. (2010) in Figure 5. As illustrated in this figure, the mean values of posteriors using the proposed method are close to the results by Zhang et al. (2010). However, it is seen that the standard deviation is decreased in all the variables using the proposed method, which yields lower coefficients of variation in the updated soil properties. This decrease

indicates that the proposed method is more reliable and its difference with RSM for multi-dimensional systems with multiple responses can be significant.

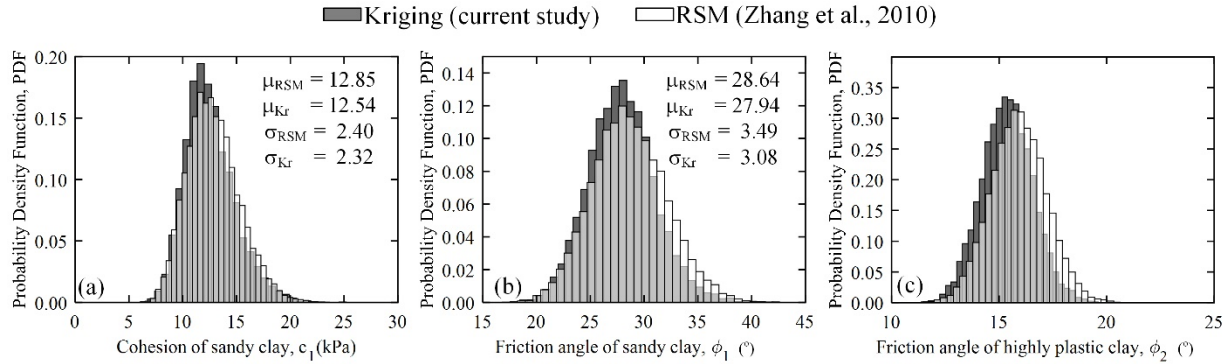


Figure 5. Comparison between the posteriors using RSM and proposed methodology (a) cohesion of sandy clay, c_1 ; (b) friction angle of sandy clay, ϕ_1 ; (c) friction angle of highly plastic clay, ϕ_2 .

3.2 Example 2: Consolidation of a clay layer

This example, adapted from Kelly and Huang (2015), is a one-dimensional consolidation soil problem with an undrained and a drained boundary at the bottom and top, respectively, of a layer of soft clay soil (Figure 6). In this example, two types of physical quantity including settlements and excess pore pressures at the base of the soft clay layer are considered to update four involved system variables $\theta = \{m_v, H, \gamma_f, C_v\}$. Here, m_v is the coefficient of volume compressibility, H is the thickness of the soil layer, γ_f is the unit weight of the fill, and C_v is the coefficient of consolidation. In this example, the real material properties are taken from Kelly and Huang (2015) and are assumed to be true values in Table 3. Moreover, similar to Kelly and Huang (2015) here we deviate the mean values of prior probabilistic models from the true values in order to evaluate the accuracy of the method. The prior probabilistic models are presented in Table 3.

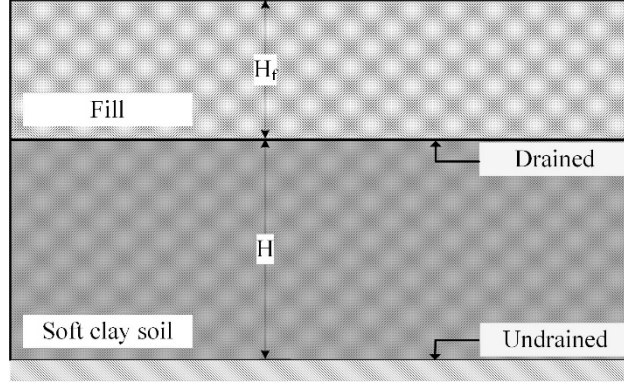


Figure 6. One-dimensional consolidation soil system with an undrained and a drained boundary.

Table 3: True material properties and prior distributions in Example 2.

Soil Parameters	Prior distributions			True values
	Distribution	Mean (μ)	C.O.V	
Coefficient of volume compressibility, m_v (1/kPa)	Lognormal	0.001	0.40	0.0014
Thickness of the soil layer, H (m)	Lognormal	5.0	0.10	5.50
Unit weight of fill, γ_f (kN/m ³)	Lognormal	20.0	0.10	22.0
Coefficient of consolidation, C_v (m ² /year)	Lognormal	40.0	1.00	80.0

Based on Atkinson (2007), the analytical solution for settlements at the bottom of the elastic soft clay layer is as,

$$s(\theta) = m_v H \gamma_f H_f \left(1 - \sum_{q=0}^{\infty} \frac{2}{\left(\frac{\pi}{2}(2q+1)\right)^2} e^{-\left(\frac{\pi}{2}(2q+1)\right)^2 T_v} \right) \quad (18)$$

where H_f is fill thickness, and T_v is the time factor determined as $T_v = C_v t / H_{dr}^2$ where t is time, and H_{dr} is the average longest drain path during consolidation, which in this example $H_{dr} = H$, since the consolidation is only in one dimension. Similar to Kelly and Huang (2015), in this study we consider $q = 9$. The second response that we are interested in capturing during the consolidation process is the excess pore pressures at the base of the soft clay layer, which can be calculated as follows (Atkinson, 2007),

$$\begin{aligned} u(\theta) &= \gamma_f H_f & T_v &< \frac{1}{12} \\ u(\theta) &= \gamma_f H_f \exp\left(\frac{1}{4} - 3T_v\right) & T_v &\geq \frac{1}{12} \end{aligned} \quad (19)$$

Using a fill thickness of 3^m and the true values of material properties from Table 3 in Equations (18) and (19), the artificial measurements are calculated for 21 time instances (Figure 7).

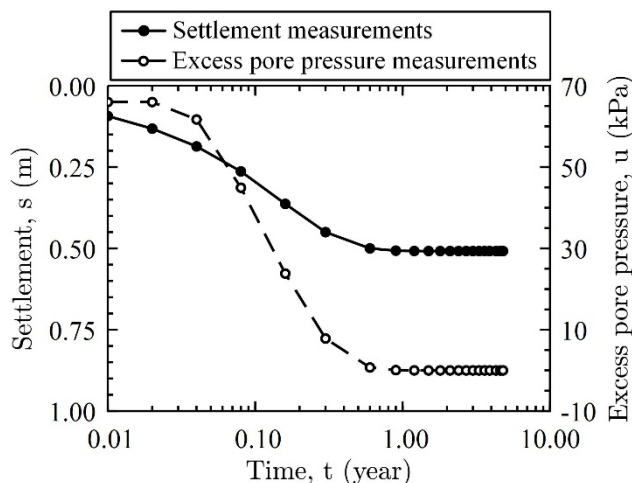


Figure 7. Artificial measurements based on true values of material properties.

To evaluate the impact of including multiple types of responses in the calibration process, a total of six cases are considered for determining the posterior responses of pore pressure and settlement. In the first three cases, the posterior responses for both pore pressure and settlement are respectively estimated based on considering only measured settlement at first time instance ($n = 1, S \text{ only}$), only measured pore pressure at first time instance ($n = 1, P \text{ only}$), and both measurements of settlement and pore pressure at first time instance ($n = 1, P \text{ and } S$). The second three cases are respectively $S \text{ only}$, $P \text{ only}$, and $P \text{ and } S$ with the considered responses from first to third time instance ($n = 3$). Variations in posterior settlement over all time instances using $n = 1$ and $n = 3$ are shown in Figures 8a and 8b, respectively, while Figures 8c and 8d illustrate variations in posterior pore pressures over all time instances for the cases of $n = 1$ and $n = 3$. Regarding the posterior settlements (Figures 8a and 8b), regardless of the value of n , the posterior responses are closer to the measurements when both responses, $P \text{ and } S$, are considered in the calibration process. Similar to the settlements, in the case of posterior pore pressures (Figures 8c and 8d), using both types of responses yield a better agreement with the measurements. However, when using only

one type of response, using the one that corresponds to the response of interest yields a better agreement compared with using other responses. As for considering more responses over time, rather than only the responses at the first time instance, a comparison can be made between Figures 8a, 8c and Figures 8b, 8d. From this comparison, it is evident that for both pore pressure and settlement $n = 3$ yields a better agreement with measured responses compared with $n = 1$. The outcomes that are achieved by using the analytical formulae are consistent with Kelly and Huang (2015). In the following, we evaluate the impact of surrogate model type and number of training samples (n_{tr}) on the posterior distributions of involved system variables.

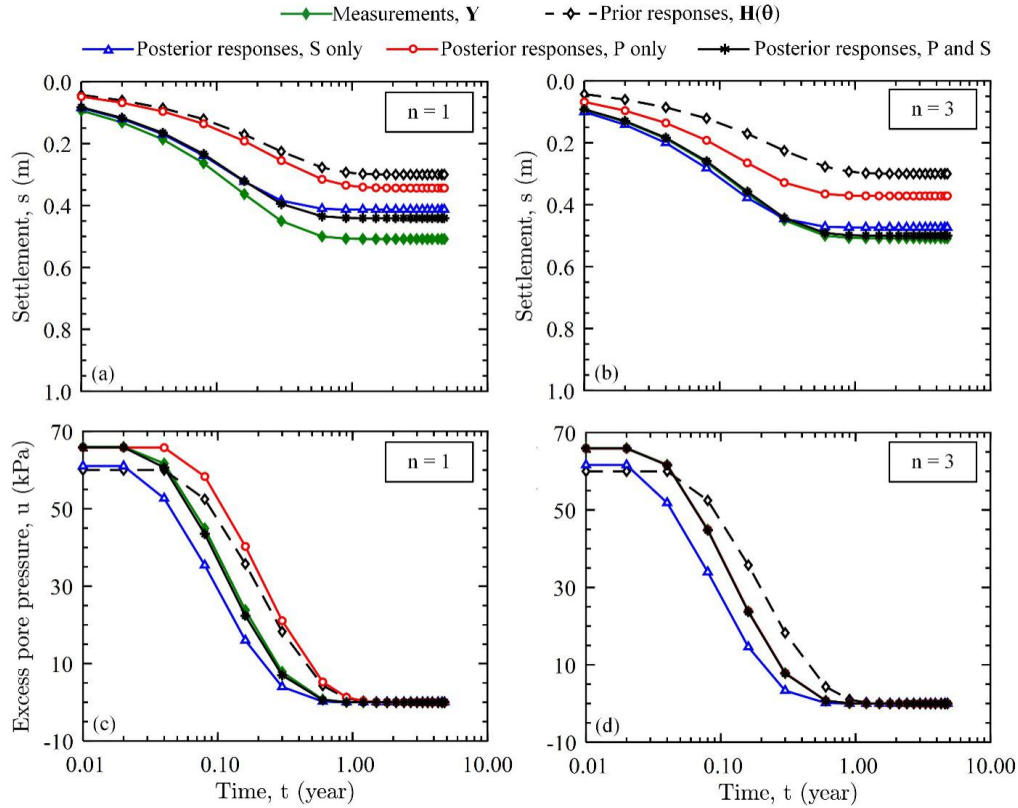


Figure 8. Impact of types and number of responses on the posterior results.

Based on the results drawn from Figure 8, we consider the case P and S for $n = 3$ and implement different types of surrogate models to evaluate the effects of surrogate models on the posterior responses. For this example, in addition to RSM, three types of surrogate models are considered to replace the simulation responses. In the first type of surrogate model, pore pressure

442 and settlement responses are considered independently from first time instance to third time
 443 instance (Surrogate 1). However, the second type of surrogate model consists of bi-variate
 444 surrogate at each time instance, which are independent of each other (Surrogate 2). In this type of
 445 surrogate model temporal correlations between responses are neglected. Finally, a multivariate
 446 surrogate model is considered for all the responses that incorporates spatiotemporal correlations
 447 of responses (Surrogate 3). The covariance matrix of $\delta(\boldsymbol{\theta})$ ($\boldsymbol{\Sigma}_\delta$) for Surrogate 1, Surrogate 2, and
 448 Surrogate 3 are in the format presented in Figure 1(c), 1(b), and 1(a), respectively. Each of these
 449 types of surrogate models are constructed with 15, 25, 40, and 80 training samples. The posterior
 450 results are presented for settlements (Figures 9a, 9b, 9c, and 9d) and pore pressures (Figures 9e
 451 and 9f). Regardless of the number of initial training samples, RSM cannot provide a sufficiently
 452 accurate surrogate model. This can be attributed to the fact that RSM assumes a homoscedastic
 453 uncertainty over the predictions. Although Surrogate 2 can provide a significantly better posterior
 454 response compared to Surrogate 1, the difference between Surrogate 2 and Surrogate 3 is not
 455 considerable. Therefore, one can conclude that the dependency between different types of
 456 responses at each snapshot of time is important in the results. Moreover, considering the
 457 dependencies between the responses in different snapshots of time can yield results that are closer
 458 to measurements; however, the gained improvements are not significant. Regarding the number of
 459 training samples, it is evident from Figure 9(a) and 9(e) that 15 number of training samples are not
 460 enough for this four dimensional example. As the number of training samples increases, the
 461 surrogate models yield closer results to the analytical formulae. Although 40 number of training
 462 samples are enough to reach acceptable agreement for the posterior responses of excess pore
 463 pressure, in order to achieve acceptable results for the posterior settlements at least 80 training
 464 samples are necessary. This can be attributed to the fact that all four random variables are involved

in settlement analysis, while three random variables impact excess pore pressures. Moreover, the extent of nonlinearity of the settlement relationship in Equation (18) is higher than that for pore pressure in Equation (19).

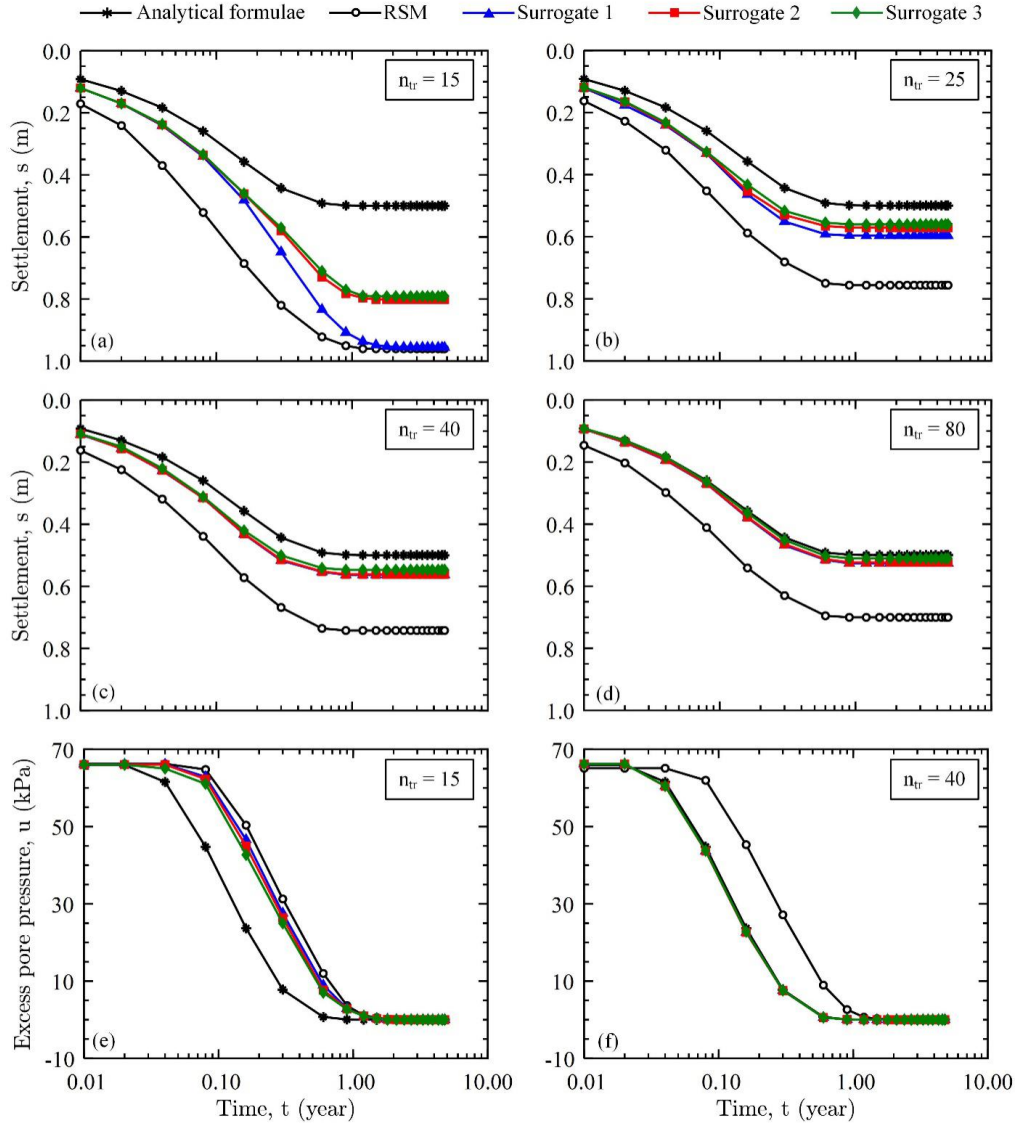


Figure 9. Impact of surrogate types and number of training points on the posterior results.

In Figure 9, the posterior responses are the responses obtained when the mean values of the posterior distributions are used. In other words, the responses in those results are $\mathbf{H}(\mathbf{E}[\boldsymbol{\theta}])$, which are not necessarily equal to $\mathbf{E}(\mathbf{H}[\boldsymbol{\theta}])$, which corresponds to our desired responses. Since analytical formulae have been used for the responses in this example, it is not time consuming to

obtain $E(H[\theta])$. Therefore, we generated 100 random samples from the posterior distributions of settlement and evaluated the true responses of each one of them using the analytical formulae. This approach yields a distribution of responses for any point from time instance 1 to 21. Figure 10 shows the resultant mean values along with the shaded region that indicates the distribution of responses. This figure confirms that $E(H[\theta])$ is not in general equal to $H(E[\theta])$. Moreover, it indicates that not only 80 training samples improve the mean values of the posterior responses but also they reduce the uncertainty in the estimated responses.

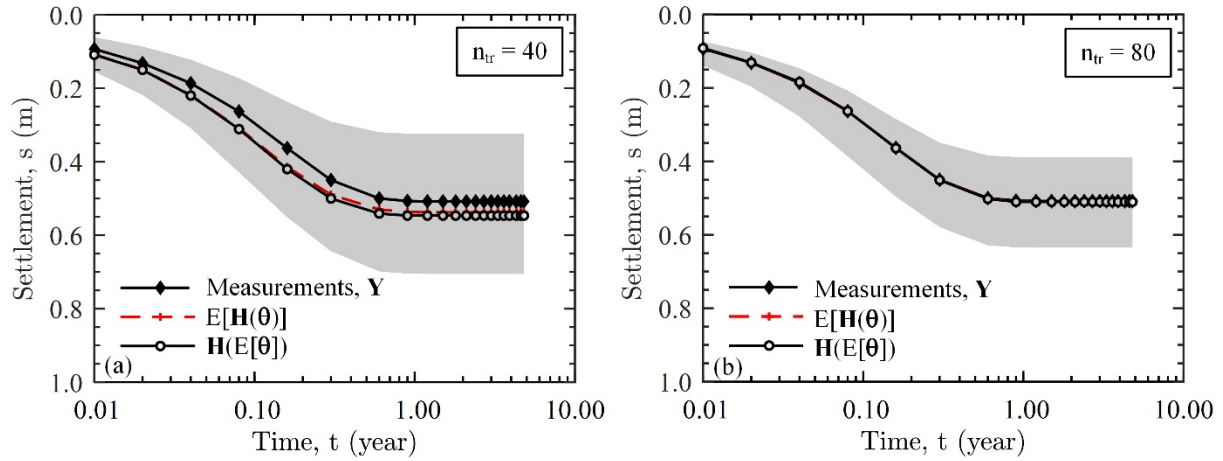


Figure 10. Comparisons between $E(H[\theta])$ and $H(E[\theta])$ for different numbers of training points; (a) $n_{tr} = 40$, (b) $n_{tr} = 80$.

3.3 Example 3: Water rise in a levee

A simple simulation of a symmetric levee located in a storm prone zone is used as the third example to illustrate the capabilities of the proposed method in a more realistic application. The levee is 8^m in height, with a 3^m concrete floodwall and an embedded sheet pile and has a 12^m wide horizontal crest; the landward and seaward slopes both have grades of 1:1. It is assumed that the levee materials are built on a sand foundation material (Figure 11a). The model geometry is simplified for the purpose of this paper; however, this approach can be extended to a real levee system with more soil layers and complicated internal structures. In this example after the initial steady-state

conditions are reached in the levee system (where the water elevation seaward side, $H_0 = 4^m$); the seaward water level is raised in 24 hours by 4^m ($H_{24} = 8^m$) in a linear manner (Figure 11b). As the water rises, a fully coupled fluid-mechanical analysis is performed using FLAC3D, and the corresponding excess pore pressures at two locations (PPT 1 and PPT 2) and horizontal top wall deflection are monitored.

In this example, the levee body is assumed to be constructed from medium sand. The parameters of interest in the coupled fluid-mechanical analysis are elastic modulus of dam (E_d), soil density of dam (ρ_d), cohesion of dam (c_d), friction angle of dam (ϕ_d), elastic modulus of foundation (E_f), soil density of foundation (ρ_f), soil permeability (K), and Biot modulus (M). The statistical properties of these parameters are shown, along with the true values (assumption by the authors), in Table 5. The process of selecting the true values in this example is similar to that in Example 2, which is explained in detail there.

Table 5: True material properties and prior distributions in Example 3.

Soil Parameters	Prior distributions			True values
	Distribution	Mean (μ)	C.O.V	
Elastic modulus of dam, E_d (Pa)	Lognormal	2.0e8	0.30	2.6e8
Soil density of dam, ρ_d (kg/m ³)	Lognormal	1500	0.05	1700
Cohesion of dam, c_d (Pa)	Lognormal	3.0e4	0.30	4.0e4
Friction angle of dam, ϕ_d (degree)	Lognormal	35	0.10	40
Elastic modulus of foundation, E_f (Pa)	Lognormal	1.2e9	0.30	1.3e9
Soil density of foundation, ρ_f (kg/m ³)	Lognormal	1900	0.05	2100
Soil permeability, K (m/s)	Lognormal	1e-6	2.50	1.0e-8
Biot modulus, M (Pa)	Lognormal	0.4e10	0.20	1.0e10

Using the true values, the artificial measurements of excess pore pressures in locations 1 and 2 are shown in Figure 11c, and for horizontal top wall deflection in Figure 11d.

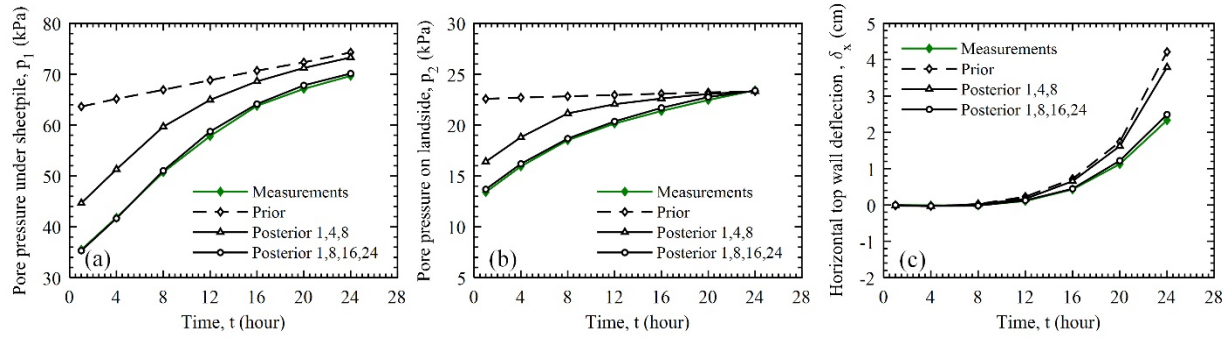


Figure 12. Variations of measurements, prior responses, posterior responses in scenario 1 and 2 versus time for (a) PPT1, (b) PPT2, and (c) Horizontal top wall deflection.

As illustrated in this figure, the prior responses and measurements are far from each other. All of the three responses have two different trends before and after $t = 8$ hours. Therefore, although the posterior responses produced by scenario 1 are closer to the measurements, compared with the prior responses, this scenario cannot provide posterior results that are acceptably close to the measurements. On the other hand, since scenario 2 uses the information in the whole window of time, it produces high agreement with the measurements. The updated probabilistic properties based on Scenario 2 are the most probable values, and they are close to the true values in this artificial example.

4. Conclusion

This paper proposed a probabilistic calibration method using multivariate surrogate modeling for multiple types of responses measured at different times and locations. This goal has been achieved by proposing a general, inclusive structure for the likelihood function and integrating it with multivariate Kriging modeling with non-homoscedastic model uncertainty over the input variable space. The proposed likelihood function also accounts for the probabilistic associations among the responses of different types to most effectively use the measurement data to update model parameters. Details of the proposed method including the general likelihood structure and the Bayesian framework, MCMC sampling, and multivariate surrogate modeling are elaborated in the

paper. The main advantages of the proposed method are: (i) direct applicability to calibration of a wide range of models for built and natural systems, (ii) suitability for high dimensional problems with different types of responses measured at different times and locations, and (iii) very high computational efficiency especially for calibration of demanding computational models.

The proposed method was applied to three examples, two from previous studies and one new example that contains highly uncertain geo-mechanical properties. These examples captured different geological and geotechnical phenomena and had different configurations and types of random variables. In each example, different goals were pursued, which included: (i) assessing the impact of non-uniform uncertainty over homoscedastic uncertainty in the prediction error, (ii) comparing the prediction accuracy by incorporating field measurement responses independent in both time and type; independent in time but dependent in type; and dependent in both time and type, and (iii) evaluating the performance of the proposed method for high dimensional systems, which is often the case in geological and geotechnical systems. The obtained results demonstrate that the highest prediction accuracy is achieved when the dependencies in time and type are incorporated in the Bayesian updating framework. Because the framework that is introduced in this study have the ability to improve the prediction accuracy through consideration of complex dependencies of multiple responses and heteroscedastic uncertainty in the design space, it can be highly efficient in geological and geotechnical engineering fields.

Acknowledgements

The U.S. National Science Foundation (NSF) through award CMMI-1563372 supported this work.

References

- Adhikari, S., Song, C. R., & Cheng, A. H. D. (2014). Evaluation of I-wall in New Orleans with back-calculated total stress soil parameters. *Acta Geotechnica*, 9(6), 1123-1134. <https://doi.org/10.1007/s11440-013-0264-1>.
- Atkinson, J. (2007). *The mechanics of soils and foundations*. CRC Press.
- Brooks, S. P., & Roberts, G. O. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing*, 8(4), 319-335.
- 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\(985\)](https://doi.org/10.1061/(ASCE)1090-0241(2004)130:10(985)).
- Duncan, J. M. (1999). Use of back analysis to reduce slope failure risk. *Civil engineering practice*, 14(1), 75-91.
- Ering, P., & Babu, G. S. (2016). Probabilistic back analysis of rainfall induced landslide-A case study of Malin landslide, India. *Engineering geology*, 208, 154-164. <https://doi.org/10.1016/j.enggeo.2016.05.002>.
- Fricke, T., Oakley, J., & Urban, N. M. (2010). Multivariate emulators with nonseparable covariance structures. URL http://www.mucm.ac.uk/Pages/Dissemination/Dissemination_Papers_Technical.html.
- Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410), 398-409.
- Gelman, B.A., Carlin, B.P., Stern, H.S. & Rubin, D.B. (2013). *Bayesian Data Analysis*, 3rd Ed., Chapman & Hall, London, United Kingdom. <https://doi.org/10.1201/b16018>.

587 Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). Markov chain Monte Carlo in practice.
 588 Chapman and Hall/CRC. <https://doi.org/10.1201/b14835>.
 589 Itasca Consulting Group, Inc. (2017) FLAC3D — Fast Lagrangian Analysis of Continua in Three-
 590 Dimensions, Ver. 6.0. Minneapolis: Itasca.
 591 Jiang, S. H., Papaioannou, I., & Straub, D. (2018). Bayesian updating of slope reliability in
 592 spatially variable soils with in-situ measurements. *Engineering Geology*, 239, 310-320.
 593 <https://doi.org/10.1016/j.enggeo.2018.03.021>.
 594 Juang, C. H., Luo, Z., Atamturktur, S., & Huang, H. (2012). Bayesian updating of soil parameters
 595 for braced excavations using field observations. *Journal of Geotechnical and*
 596 *Geoenvironmental Engineering*, 139(3), 395-406. [https://doi.org/10.1061/\(ASCE\)GT.1943-](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000782)
 597 [5606.0000782](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000782).
 598 Kelly, R., & Huang, J. (2015). Bayesian updating for one-dimensional consolidation
 599 measurements. *Canadian Geotechnical Journal*, 52(9), 1318-1330. [https://doi.org/10.1139/cgj-](https://doi.org/10.1139/cgj-2014-0338)
 600 [2014-0338](https://doi.org/10.1139/cgj-2014-0338).
 601 Khosravi, A., Rahimi, M., Gheibi, A., & Mahdi Shahrabi, M. (2017). Impact of Plastic
 602 Compression on the Small Strain Shear Modulus of Unsaturated Silts. *International Journal of*
 603 *Geomechanics*, 18(2), 04017138. [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0001031](https://doi.org/10.1061/(ASCE)GM.1943-5622.0001031).
 604 Li, D. Q., Jiang, S. H., Cao, Z. J., Zhou, W., Zhou, C. B., & Zhang, L. M. (2015). A multiple
 605 response-surface method for slope reliability analysis considering spatial variability of soil
 606 properties. *Engineering Geology*, 187, 60-72. <https://doi.org/10.1016/j.enggeo.2014.12.003>.
 607 Li, S., Zhao, H., Ru, Z., & Sun, Q. (2016a). Probabilistic back analysis based on Bayesian and
 608 multi-output support vector machine for a high cut rock slope. *Engineering geology*, 203, 178-
 609 190. <https://doi.org/10.1016/j.enggeo.2015.11.004>.

610 Li, X. Y., Zhang, L. M., Jiang, S. H., Li, D. Q., & Zhou, C. B. (2016b). Assessment of slope
611 stability in the monitoring parameter space. *Journal of Geotechnical and Geoenvironmental*
612 *Engineering*, 142(7), 04016029. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001490](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001490).

613 Li, X. Y., Zhang, L. M., & Jiang, S. H. (2016c). Updating performance of high rock slopes by
614 combining incremental time-series monitoring data and three-dimensional numerical analysis.
615 *International Journal of Rock Mechanics and Mining Sciences*, 83, 252-261.
616 <https://doi.org/10.1016/j.ijrmms.2014.09.011>.

617 Matthies, H. G., Brenner, C. E., Bucher, C. G., & Soares, C. G. (1997). Uncertainties in
618 probabilistic numerical analysis of structures and solids-stochastic finite elements. *Structural*
619 *safety*, 19(3), 283-336. [https://doi.org/10.1016/S0167-4730\(97\)00013-1](https://doi.org/10.1016/S0167-4730(97)00013-1).

620 Miro, S., König, M., Hartmann, D., & Schanz, T. (2015). A probabilistic analysis of subsoil
621 parameters uncertainty impacts on tunnel-induced ground movements with a back-analysis
622 study. *Computers and Geotechnics*, 68, 38-53. <https://doi.org/10.1016/j.compgeo.2015.03.012>

623 Qi, X. H., & Zhou, W. H. (2017). An efficient probabilistic back-analysis method for braced
624 excavations using wall deflection data at multiple points. *Computers and Geotechnics*, 85,
625 186-198. <https://doi.org/10.1016/j.compgeo.2016.12.032>.

626 Rahimi, M., Wang, Z., Shafieezadeh, A., Wood, D., & Kubatko, E. J. (Unpublished results).
627 Exploring Passive and Active Meta-Modeling-Based Reliability Analysis Methods for Soil
628 Slopes. *International Journal of Geomechanics*.

629 Sun, Y., Huang, J., Jin, W., Sloan, S. W., & Jiang, Q. (2019). Bayesian updating for progressive
630 excavation of high rock slopes using multi-type monitoring data. *Engineering Geology*.
631 <https://doi.org/10.1016/j.enggeo.2019.02.013>.

- Sun, Y., Jiang, Q., Yin, T., & Zhou, C. (2018). A back-analysis method using an intelligent multi-objective optimization for predicting slope deformation induced by excavation. *Engineering Geology*, 239, 214-228. <https://doi.org/10.1016/j.enggeo.2018.03.019>.
- Svenson, J. (2011). Computer experiments: Multiobjective optimization and sensitivity analysis (Doctoral dissertation, The Ohio State University).
- Wang, L., Hwang, J. H., Luo, Z., Juang, C. H., & Xiao, J. (2013). Probabilistic back analysis of slope failure—A case study in Taiwan. *Computers and Geotechnics*, 51, 12-23. <https://doi.org/10.1016/j.compgeo.2013.01.008>.
- Wang, Y., & Akeju, O. V. (2016). Quantifying the cross-correlation between effective cohesion and friction angle of soil from limited site-specific data. *Soils and Foundations*, 56(6), 1055-1070. <https://doi.org/10.1016/j.sandf.2016.11.009>.
- Wang, H. J., Xiao, T., Li, X. Y., Zhang, L. L., & Zhang, L. M. (2019). A novel physically-based model for updating landslide susceptibility. *Engineering Geology*. <https://doi.org/10.1016/j.enggeo.2019.02.004>.
- Yang, H. Q., Zhang, L., & Li, D. Q. (2018). Efficient method for probabilistic estimation of spatially varied hydraulic properties in a soil slope based on field responses: A Bayesian approach. *Computers and Geotechnics*, 102, 262-272. <https://doi.org/10.1016/j.compgeo.2017.11.012>.
- 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. <https://doi.org/10.1016/j.compgeo.2010.07.009>.
- Zhang, L., Wu, F., Zheng, Y., Chen, L., Zhang, J., & Li, X. (2018). Probabilistic calibration of a coupled hydro-mechanical slope stability model with integration of multiple observations.

655 Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 12(3),
656 169-182. <https://doi.org/10.1080/17499518.2018.1440317>.
657 Zhou, X., Chen, J., Chen, Y., Song, S., Shi, M., & Zhan, J. (2017). Bayesian-based probabilistic
658 kinematic analysis of discontinuity-controlled rock slope instabilities. Bulletin of engineering
659 geology and the environment, 76(4), 1249-1262.