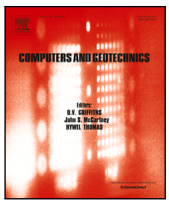
Data assimilation for Bayesian updating of predicted embankment response using monitoring data

Computers and Geotechnics 165 (2024) 105936

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo



Research paper

Data assimilation for Bayesian updating of predicted embankment response using monitoring data

Amardeep Amavasai [a,*], Hossein Tahershamsi [b], Tara Wood [c], Jelke Dijkstra [a]

[a] Chalmers University of Technology, Division of Geo-engineering, SE-41296 Gothenburg, Sweden
[b] ELU Konsult AB, Västra Hamngatan 14, SE-41117 Gothenburg, Sweden
[c] Ramboll Sverige AB, Vädursgatan 6, SE-41250 Gothenburg, Sweden

ARTICLE INFO

Dataset link: https://github.com/amaran1988/DA-PLAXIS2D.git

Keywords:
Data assimilation
Embankment
Soft soil
Plaxis
Monitoring data

ABSTRACT

Data Assimilation (DA) algorithms has been successfully employed in geotechnical problems to jointly estimate the state of the system and model parameters, however, the impact of the field monitoring setup on the performance of DA is often overlooked. In this paper, the impact of the field monitoring setup on the performance of DA is studied. The Ensemble Kalman Filter is used as the DA algorithm as part of a synthetic experiment which includes a fully coupled hydromechanical numerical model of an embankment constructed on soft ground. The results of the assimilated parameters show different rate of convergence toward their synthetic true value which corroborates well with the results of the global sensitivity analysis performed in this study. In order to investigate the difference in influence between the quantity and type of measurement, different monitoring strategies were chosen in this study. The results indicate that the effective friction angle and Poisson's ratio are better estimated when the horizontal displacement is included along with the vertical displacement in the observation space of the DA procedure. Finally, the strong correlation between observation type and parameter convergence is independent of the type of initial prior knowledge, but strongly depends on the measurement location.

1. Introduction

Efficient design for geotechnical infrastructure in terms of serviceability limit state depends greatly on accurately forecasting the current and future response. Deterministic forward models are commonly used to predict the behaviour of a geotechnical system. The system response, however, is susceptible to various uncertainties that stem from limited knowledge and/or poor understanding of the physical phenomena. Improving the performance of these models can be achieved by adjusting their parameters to values that are more close to the system response, preferably through an inverse method, by utilising measurements, which can help to estimate the poorly known model parameters.

With recent technological advancements, projects are now equipped with tools for real-time monitoring, leading to new possibilities for validating the ever-growing optimisation techniques for inverse analysis and subsequent uncertainty reduction in the model forecast. However, dealing with monitoring data can be challenging due to the quality of the acquired dataset leading to uncertainties. In order to accurately determine the probable range of behaviours with a practical margin of error, it is crucial to consider the uncertainties arising from both the monitoring data and the numerical simulation.

In recent times, there has been a growing interest in employing the Bayesian statistical framework to address the task of jointly estimating the state and parameters using monitoring data (Wu et al., 2007; Zhang et al., 2010; Juang et al., 2013). Recent developments in other scientific domains have shown the efficacy of a powerful method known as Data Assimilation (DA), which rigorously incorporates observations into numerical forecasting models accounting for uncertainties. The Ensemble Kalman Filter (EnKF), a sequential DA technique that incorporates data into the model prediction as they become available (*i.e.*, sequentially), is one of the many techniques in the DA framework that has lately gained popularity in geotechnical applications (Liu et al., 2018; Tao et al., 2020; Mohsan et al., 2021; Tao et al., 2021, 2022). The probability distributions are represented by their sample realisation, referred to as ensembles, and hence can easily deal with nonlinear state equations making it feasible for geotechnical applications. This prior work shows that EnKF is an efficient Bayesian statistical method and has been successfully employed in geotechnical problems, however, it is still not fully attractive for general geotechnical practice

* Corresponding author.
E-mail addresses: amardeep@chalmers.se (A. Amavasai), hossein.tahershamsi@elu.se (H. Tahershamsi), tara.wood@ramboll.se (T. Wood), jelke.dijkstra@chalmers.se (J. Dijkstra).

https://doi.org/10.1016/j.compgeo.2023.105936
Received 5 August 2023; Received in revised form 9 November 2023; Accepted 10 November 2023
Available online 16 November 2023

since many issues remain unanswered. For example, the impact of the field monitoring setup on the performance of DA, particularly in geotechnical engineering, is frequently underestimated and overlooked, with recommendations typically relying on experience. While empirical decisions are generally not incorrect, the measurement arrangement, *i*.e., the quantity, type and location, exert a major influence on the results of the back-analysis. This is because not all model parameters have the same influence on the state response and the sensitivity of those parameters vary in the spatiotemporal domain. The variation in sensitivity can result in different results in terms of accuracy and precision for each parameter when employing Data Assimilation with different monitoring configurations. In view of this, it makes sense to pose the following questions:

1. Does an increased quantity of measurements give a "significant" improvement in the accuracy and precision of the parameters and prediction of the system performance?
2. What type, or combination, of measurements are needed for an accurate prediction of the parameters?
3. How much influence does the spatiotemporal sensitivity of the parameters have on the convergence of the parameters during Data Assimilation ?
4. Does the type of prior information included in the Data Assimilation procedure influence the results?
5. What are the effects of the measurement location on the performance of Data Assimilation?

This paper aims to answer the above questions through a synthetic numerical experiment of an embankment constructed on soft soil. In this study, the constitutive behaviour of the soil is modelled using the Soft Soil model, available in general geotechnical Finite Element code PLAXIS (Plaxis, 2015). The chosen geometry of the embankment, the soil type and the aforementioned constitutive model are ideal in nature, so it should be noted that the conclusions made for the above questions may not be generic but offer a detailed insight and a way forward for researchers to pursue the above questions in other scenarios. In this study the sensitivity of the parameters is calculated using the factorial analysis (Tahershamsi and Dijkstra, 2022) due to its computational efficiency.

This paper is organised as follows: First, an overview of the concepts and methods of Data Assimilation are presented, and the theory underpinning the Ensemble Kalman Filter (EnKF) is elaborated. Then the methodology where the process of integrating Data Assimilation in a Finite Element code is explained along with the details of the synthetic experiment and Global Sensitivity Analysis (GSA) used in this study. This is followed by an initial exercise where the state of the system and the parameter vector are jointly back-calculated from a pre-defined search space using Data Assimilation. The primary objective here lies in the statistical assessment of the identified parameters and the state of the system for a given monitoring setup. Then the subsequent sections deal with the aforementioned questions where the effect of sensor strategy and type of prior distribution on the performance of the Data Assimilation is evaluated. Finally, the paper concludes with the results of the findings and future research steps.

2. Data assimilation

Data Assimilation (DA) is a versatile approach to estimate the state of a dynamic system by combining observational data with a numerical model forecast of the system. Considering a time window *i*.e., $t \in [0,T]$, the evolution of the state in the spatiotemporal domain, $f(t,x)$, is dictated by the forward model $\mathbb{F} : f(t,x) \rightarrow f(t+\Delta t,x),\ \forall (t,x) \in [0,T]\times\Omega$ as shown in Eq. (1) where the state of the system, represented by $\psi_k \in \mathbb{R}^m$, is evolved using the forward dynamics:

$$\psi_k = \begin{pmatrix} u_k \\ p_k \end{pmatrix} \in \mathbb{R}^m$$
$$\psi_k = F(\psi_{k-1}) + q_{k-1} \tag{1}$$

where, u_k and p_k are the nodal displacement and excess porewater pressure vector respectively for time-step k of the discretised system and $q_k \sim N(0,\sigma_q)$ is the process noise vector for the modelling errors. $\psi_k \in \mathbb{R}^m$ is often called the 'model space' or 'prediction space'. The true behaviour of the system is observed via a set of discrete points in the domain (representing instruments in the ground) and this is often termed as the 'observation space' modelled by $y_k \in \mathbb{R}^n$. The mapping from the model space to the observation space is performed via the operator $\mathbb{H} : f(t,x) \rightarrow g(t,x),\ \forall (t,x) \in [0,T]\times\Omega$:

$$y_k = H_k^{\psi}(\psi_k) + v_k \tag{2}$$

where, $v_k \sim N(0,\sigma_v)$ is the noise corrupting the measurement. The error covariance matrix for the process noise is given as $E[q_k\ q_k^T] \rightarrow Q_k$ and for the observation error the covariance matrix is defined as $E[v_k\ v_k^T] \rightarrow R_k$.

The state space, $\psi \in \mathbb{R}^m$, needs to be augmented to form a combined state-parameter space, to simultaneously estimate the model parameter set, $\theta \in \mathbb{R}^p$, along with the evolving model state. The joint space x_k enables to simultaneously update both the model parameters and the state variables during the assimilation process (Iglesias et al., 2013; Bocquet and Sakov, 2013):

$$x_k = \begin{pmatrix} \psi_k \\ \theta_k \end{pmatrix} \in \mathbb{R}^{m+p} \tag{3}$$

The augmented system state can be integrated with the governing equations for the evolution of the model state, following the usual approach. It is important to note, however, that the observation operator needs to be augmented as well, since the parameters are not part of the observation space:

$$H_k = \begin{pmatrix} H_k^{\psi} & 0 \end{pmatrix} \in \mathbb{R}^{n\times(m+p)} \tag{4}$$

The augmented state vector allows for the calculation of the cross-covariance between the states and parameters. The inference of the unobserved parameter and its uncertainty for the joint state-parameter estimation relies crucially on the cross covariance matrix (see Eq. (5)). The off-diagonal elements of the (augmented) state error covariance matrix $E\ [\ (x-E[x])\ (x-E[x])^T\ ]$ pass information from the data assimilated state, to improve the estimate of the unobserved parameters:

$$E\ [\ (x-E[x])\ (x-E[x])^T\ ] = \begin{pmatrix} P_{\psi\psi} & P_{\psi\theta} \\ P_{\theta\psi} & P_{\theta\theta} \end{pmatrix} \tag{5}$$

In this study, the explicit consideration of model error is not included due to the adopted persistence model ($d\theta/dt = 0$), *i*.e., the parameter set remains constant during the model state evolution until the subsequent assimilation cycle. This is because the uncertainty in the parameters implicitly accounts for the model error, and introducing any further errors potentially diminishes the importance of previous assimilation results, thus affecting the convergence (Trudinger et al., 2008).

In this work, the overall convergence rate of the parameters is analysed by the normalised trace of the parameter covariance matrix, termed as the Uncertainty Ratio, Z_A given in Eq. (6). For assessing the rate of convergence for each individual parameter, the normalised variance of its estimate is given in terms of the Uncertainty ratio, Z, as shown below:

$$Z_A = \mathrm{Tr}\,(P_{\theta\theta})/\,\mathrm{Tr}\,(P_{\theta\theta}^0)$$
$$Z = \sigma_{\theta\theta}^2/\sigma_{\theta\theta_0}^2 \tag{6}$$

Using all the available noisy observations until time 'k' ($y_{1:k} = \{y_1, y_2 \ldots y_k\}$), the state of the system can be updated to construct the posterior density. For this, the prior distribution at time 'k' is required which, based on the Chapman–Kolmogorov equation, is obtained by projecting the previous posterior at time 'k-1' forward in time using the forward model (Chatzi and Smyth, 2009).

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})\ p(x_{k-1}|y_{1:k-1})\ dx_{k-1} \tag{7}$$

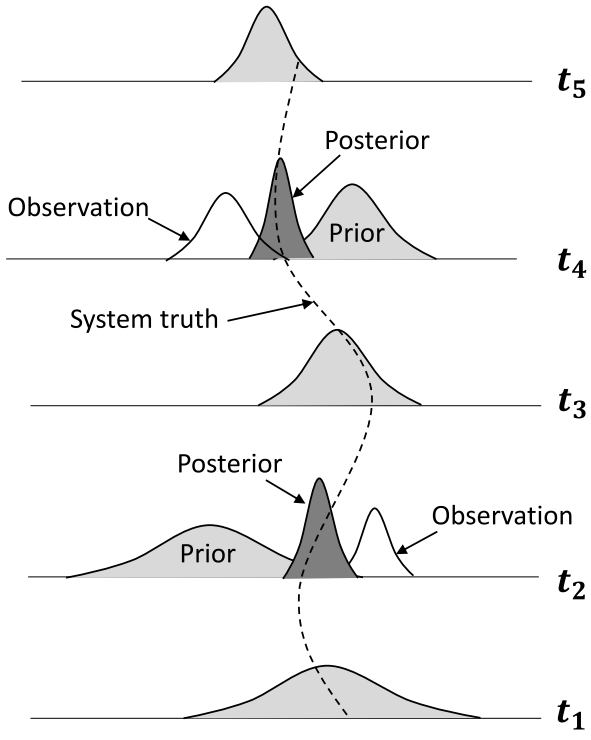


Fig. 1. Illustration of the working of sequential Data Assimilation (t_1 to t_5 are time intervals).

Subsequently the updated posterior at time 'k' can be obtained, by incorporating the latest observation likelihood:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)\ p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \tag{8}$$

The process of this sequential assimilation of observations is summarised in Fig. 1. The typical assimilation scheme consists of two main stages: (1) a prediction/forecast step that serves as the prior for that timestep, and (2) an update/analysis step that incorporates the observations to obtain the posterior. From Fig. 1, the predicted result from time t_1 to t_2 using the forward model represents the prior knowledge at t_2. The observation at time t_2 is then integrated to obtain the posterior. Using this posterior and the forward model, the state at t_3 is predicted. As there is no observation in this time period, the simulation continues. At time t_4, when the observation is made available, the prior is updated using this observation to obtain the posterior which is then propagated to time t_5. However, solving the Bayesian integral equation is usually intractable, particularly for highly nonlinear problems. Therefore, it is essential to employ algorithms that provide approximate solutions, one of those will be discussed in the following section.

2.1. Ensemble Kalman filter

The Ensemble Kalman Filter (EnKF) is a variation of the Kalman filter (Kálmán, 1960) introduced by Evensen (1994). In EnKF the statistical properties of the state variable are represented by a collection of ensemble members, thereby approximating the covariance matrix, which are evolved over time using the nonlinear dynamics of the system. Thus instead of storing the full covariance matrix, EnKF can represent the same error statistics using an appropriate set of ensembles. As the population of the ensemble increases, the error in the approximation decreases. In the strict sense, EnKF always introduces error due to its finite-size approximation of the covariance matrix. Subsequently, by assuming gaussian approximation, the ensemble set is subjected to the standard Kalman filter to compute the posterior mean and variance at each assimilation step. The Ensemble Kalman Filter (EnKF) is attractive due to its ability to effectively handle high-dimensional problems with a small ensemble size (Schillings and Stuart, 2016). Despite its use in geoscience, EnKF is still a relatively new approach in geotechnics. Recent studies have investigated the potential of EnKF in geotechnics (e.g Vardon et al., 2016; Liu et al., 2018; Mavritsakis, 2017). The workflow of EnKF is shown in Fig. 2:

The ensemble representation of the augmented state vector is expressed as:

$$x_k^N = \begin{pmatrix} \psi_k^1 & \psi_k^2 & \cdots & \psi_k^N \\ \theta_k^1 & \theta_k^2 & \cdots & \theta_k^1 \end{pmatrix} \in \mathbb{R}^{N\times(m+p)} \tag{9}$$

The mean, anomaly and the subsequent covariance matrix of the augmented ensemble forecast state vector is estimated as:

$$\begin{aligned} x_k^f &= x_k^{f,N} \cdot \mathbf{1}_N \\ \boldsymbol{X}_k^f &= \frac{1}{\sqrt{N-1}}\,(x_k^{f,N} - x_k^f) \\ P_k^{e,f} &= (\boldsymbol{X}_k^f)(\boldsymbol{X}_k^f)^T \approx P_k^f \end{aligned} \tag{10}$$

where superscript e represents the quantities estimated from ensembles with $\mathbf{1}_N$ representing the equal weight vector for calculating the mean. An ensemble of perturbed observation with covariance $\boldsymbol{R}$ is defined (Burgers et al., 1998) as follows:

$$\begin{aligned} y_k^j &= y_k^t + v_k^j \quad j = 1, 2, \ldots, N \\ \boldsymbol{Y}_o' &= \frac{1}{\sqrt{N-1}}\,\left[v_k^1, v_k^2, \ldots, v_k^N\right] \\ \boldsymbol{R}^e &= (\boldsymbol{Y}_o')(\boldsymbol{Y}_o')^T \approx \boldsymbol{R} \end{aligned} \tag{11}$$

The ensemble based Kalman gain matrix $\boldsymbol{K}^e$ is obtained from the approximate background covariance matrix and observation operator (shown below as time-independent and linear, as used in this study) and each ensemble member is updated in the analysis step:

$$\begin{aligned} \boldsymbol{K}^e &= P_k^{e,f} H^T \left[H P_k^{e,f} H^T + \boldsymbol{R}^e \right]^{-1} \\ x_{n,k}^a &= x_{n,k}^f + \boldsymbol{K}^e\,[y_{n,k} - H x_{n,k}^f], \quad 1 \le n \le N \\ P_k^{e,a} &= [\boldsymbol{I} - \boldsymbol{K}^e H]\ P_k^{e,f} \end{aligned} \tag{12}$$

According to Verlaan and Heemink (2001), EnKF can be used for strongly nonlinear problems. Even though the EnKF does not solve the Bayesian update of a non-Gaussian probability density function, the updated ensemble inherits most of the non-Gaussian properties from the forecast ensemble since only the updates defined by the right-hand side of Eq. (12) for calculating $x_{n,k}^a$ are linear. Due to this, the resampling of the posterior is avoided making EnKF computationally efficient (Carrassi et al., 2017).

3. Methodology

The Data Assimilation (DA) framework allows for the estimation of model parameters by analysing sequential measurements taken under known varying traction boundary conditions. Fig. 3 illustrates the integration of the Ensemble Kalman filter (EnKF) into a geotechnical application implemented within the commercial Finite Element code PLAXIS (Plaxis, 2015). The behaviour of the system is simulated in a controlled setting by producing synthetic measurement data with noise, specifically vertical and horizontal displacements. The Data Assimilation (DA) procedure is then implemented for the considered time-dependent geotechnical system, which involves an embankment on soft soil undergoing consolidation.

Following initialisation, the numerical forecasting model utilises the set of ensembles, each with a unique set of model parameters, to predict the geotechnical response in the time domain up to a specified time step. These ensembles represent the prior belief of the parameter values leading up to the time of the available measurement. The predicted

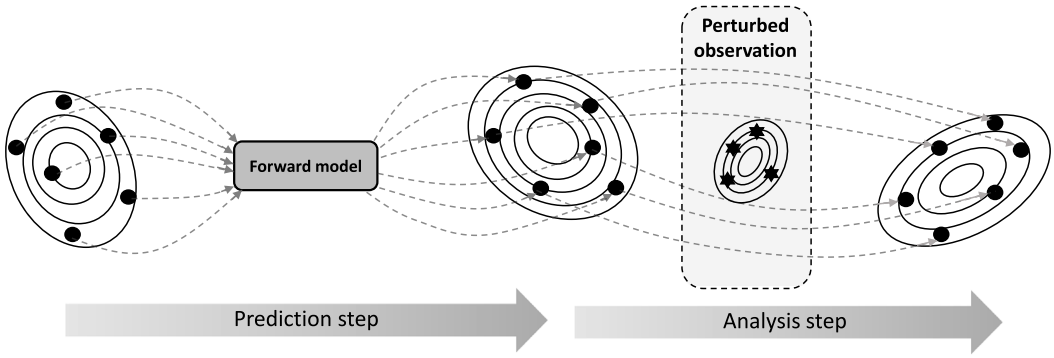


Fig. 2. Illustration of the principle of the Ensemble Kalman Filter.

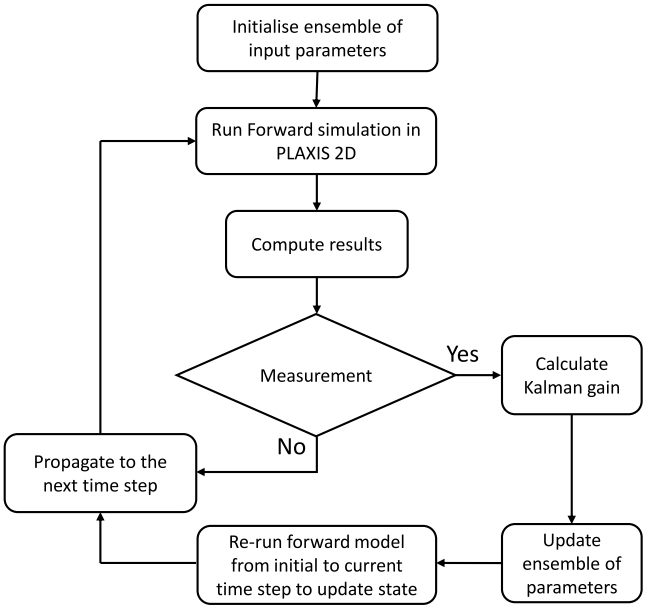


Fig. 3. Illustration of the integrated workflow of PLAXIS with Ensemble Kalman Filter.

state in the model space is subsequently transformed into the observation space, and through the use of DA, the posterior distribution of the parameter set is estimated.

The displacement at a specific time interval depend not only on the magnitude of the model parameters but also on the state variables, such as stress, strain, and porewater pressure distribution, at each time step. The evolution of these states depend on the loading history of the problem and a recursive algorithm is therefore necessary to update these states along with the model parameters at each assimilation cycle to achieve proper convergence, albeit, at an unavoidably increased computational cost (Mohsan et al., 2021). The converged parameters are assessed in terms of physical meaningfulness and subsequent model performance on future forecasts in the prediction window.

A coupled hydro-mechanical finite element model (FEM) simulation serves as the forward model. The study investigates the impact of various factors such as sensor characteristics, prior distribution type, and parameter sensitivity on the Bayesian update of embankment behaviour. The entire workflow is implemented and solved in a Python environment (Van Rossum and Drake Jr., 1995), utilising the PLAXIS Python interface to automate the finite element calculations in PLAXIS and integrating it with the DA algorithm in the same Python script. A basic version of this script is made available on GitHub (https://github.com/amaran1988/DA-PLAXIS2D.git).

4. Numerical experiment

This paper presents a numerical experiment that involves integrating the Ensemble Kalman Filter (EnKF) into a Finite Element code to update the behaviour of an embankment based on available monitoring information. The embankment is constructed in two stages. In the first stage, the embankment is constructed within a time span of 60 d to an height of 2.0 m, and allowed to consolidate for 1 year. During this consolidation period monitoring information is made available. This period ranging between 60 to 425 days after start of construction is treated as the assimilation window. The Ensemble Kalman Filter (EnKF) is then used to calibrate the parameters of the model using the monitoring data. In the second stage, *i.e.* after 425 d, the height of the embankment is further raised to 3 m over a period of 60 d. This second stage is treated as the prediction window without monitoring data in which the efficiency of the Data Assimilated/calibrated model is assessed based on the accuracy and precision of the prediction for an altered system. The embankment is 4 m wide with 6 m on either side sloping toward the toe constructed with a slope of 1:2 with an embankment material that has a unit weight of 17 kN/m^3. The discretisation of the mesh along with the dimensions of the Finite Element model is shown in Fig. 4. The numerical model is discretised to a total of 5859 15-noded triangular elements leading to a total of 47475 nodes. The behaviour of the embankment is considered deterministic and modelled as a drained material with Mohr–Coulomb constitutive model with a Young's modulus of 40000 kPa and Poisson's ratio of 0.35. The strength of the embankment material is prescribed using an effective friction angle ϕ' of 35°. The problem is simplified to a homogeneous soil profile with phreatic level at the ground surface. The hydraulic properties are assumed isotropic ($K_v = K_h$).

4.1. Constitutive model for the soft soil

The assumed soil behaviour reflect those of a slightly overconsolidated to normally consolidated soft soil deposit, therefore, the Soft Soil (SS) constitutive model has been chosen. Although inspired from the Modified Cam Clay (MCC) model (Roscoe and Burland, 1968), it cannot be classified as a Critical State Model due to the modified formulation for the yield surface. This modification enables the SS model to predict a reasonable value for the K_0, at normally consolidated region while the MCC model predicts overly high values. The SS model assumes an associated flow rule (*e.g.*, Karstunen and Amavasai, 2017; Plaxis, 2015). The trade-off for the aforementioned modification is that a separate failure condition needs to be imposed using the Mohr–Coulomb failure criterion. Also SS does not allow stress states that exceed the Mohr–Coulomb failure criterion, hence, strain softening is not permitted. However, in the current study, SS is appropriate since the soil is not heavily overconsolidated.

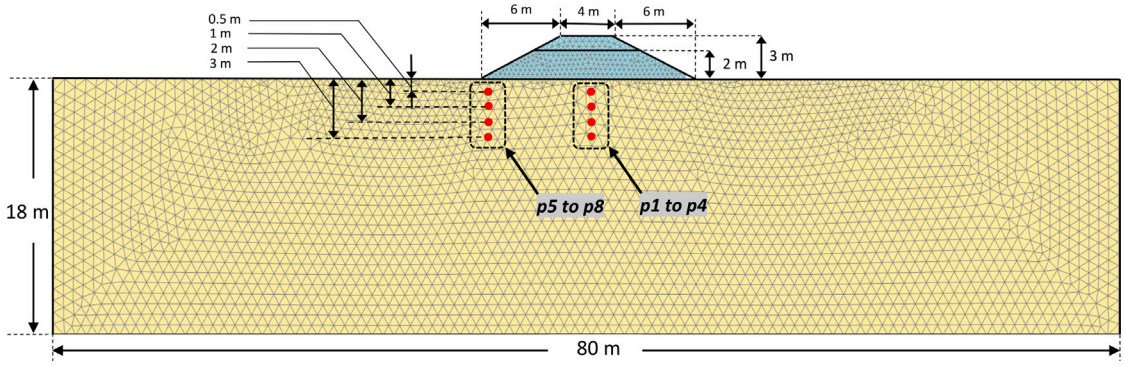


Fig. 4. Mesh discretisation and dimensions of the numerical model.

Table 1
Synthetic truth and feasible space of parameters for the Soft Soil model.

Parameter	Unit	Synthetic truth	Feasibility space
λ^*	–	0.280	0.200–0.520
κ^*	–	0.030	0.015–0.039
ν'_{ur}	–	0.200	0.150–0.380
ϕ'	degrees	27.00	24.0–59.0
POP	kPa	5.000	2.10–5.50
k	m/day	0.006	0.005–0.030

Table 2
Temporal monitoring points.

Monitoring ID	Time [days]
1	62.165
2	66.495
3	72.990
4	81.651
5	88.146
6	103.302
7	116.292
8	133.613
9	150.934
10	185.576
11	220.218
12	263.520
13	315.483
14	367.445
15	402.087
16	425.000

4.2. Generation of synthetic data

Table 1 shows the parameter values used to generate the synthetic truth from which the monitoring data will be sampled. The forward simulation is performed using the numerical model built in PLAXIS. The measured nodal displacements are obtained at specific locations beneath the embankment, at depths of 0.5 m, 1.0 m, 2.0 m, and 3.0 m. The set of measurements includes four sensors beneath the centre to monitor vertical displacement (*p1*, *p2*, *p3* and *p4*) and four sensors below the toe of the embankment to monitor horizontal displacement (*p5*, *p6*, *p7* and *p8*) as shown in Fig. 4. The objective is to show that the posterior updating is sensitive to the location of the sensors. For example, friction angle becomes more important where a shear plane is located. So, to demonstrate this in a synthetic setting, the location of the sensors were chosen accordingly. These synthetic measurements are obtained at time intervals following standard practice for settlement retrieval, with more frequent measurements at the beginning of embankment construction and progressively reduced intervals as time progresses (see Table 2). Synthetic data for time-dependent displacements are generated with Gaussian noise added to the observation data. For this study, a noise level of 5 mm is selected to generate noise around the true synthetic displacements.

4.3. Generation of ensemble population

In practical applications, numerically evolving the (nonlinear) model dynamics can be time-consuming, and there is a limitation on the size of the ensemble that can be utilised. EnKF always introduces error due to its finite-size approximation but this may become more pronounced with a reduced sample size. Drawing from a comparable case study (Mohsan et al., 2021), it is concluded that reasonable accuracy can be attained by employing an ensemble size of 50, which is adopted in this study. While the Ensemble Kalman Filter (EnKF) is typically formulated based on Gaussian assumptions, some studies in geotechnical engineering have employed prior information with non-Gaussian distributions (Tao et al., 2020, 2022). Considering this, a multivariate uniform distribution was assumed for the prior probability density function of the uncertain parameters. The distribution is independently defined within lower and upper bounds. The calculations were repeated 10 times using different initial samples, to eliminate randomness. The mean values were intentionally set differently (and far) from the synthetic true parameter values. The feasibility space, as shown in Table 1, is determined by selecting bounds that ensured a consistent coefficient of variation (COV) of 0.25 for all parameters except for the hydraulic conductivity since a larger variation (COV = 0.40 is chosen in this study) is usually observed in practice.

4.4. Global sensitivity analysis

Global Sensitivity Analysis (GSA) refers to a set of systematic approaches that aim to comprehensively investigate the sources of uncertainty in any process or system. One of the primary advantages of GSA methods is their ability to systematically explore the entire space of input parameters (Saltelli et al., 2008). Such methods are required in order to assess the effect of sensor type and their location towards the performance of a Data Assimilation procedure. Tahershamsi and Dijkstra (2022) have illustrated the advantages of using GSA methods both in the spatial and the temporal domain within geotechnical FE modelling. In addition, the authors have shown that factorial design is more economical for geotechnical models with a limited number of model parameters (<15) compared to the most commonly used Sobol method, without a significant compromise in accuracy.

In this study, a factorial design is chosen as the GSA method to evaluate the spatio-temporal sensitivity of the model parameters. A predetermined number of *levels* is chosen for each factor of interest,

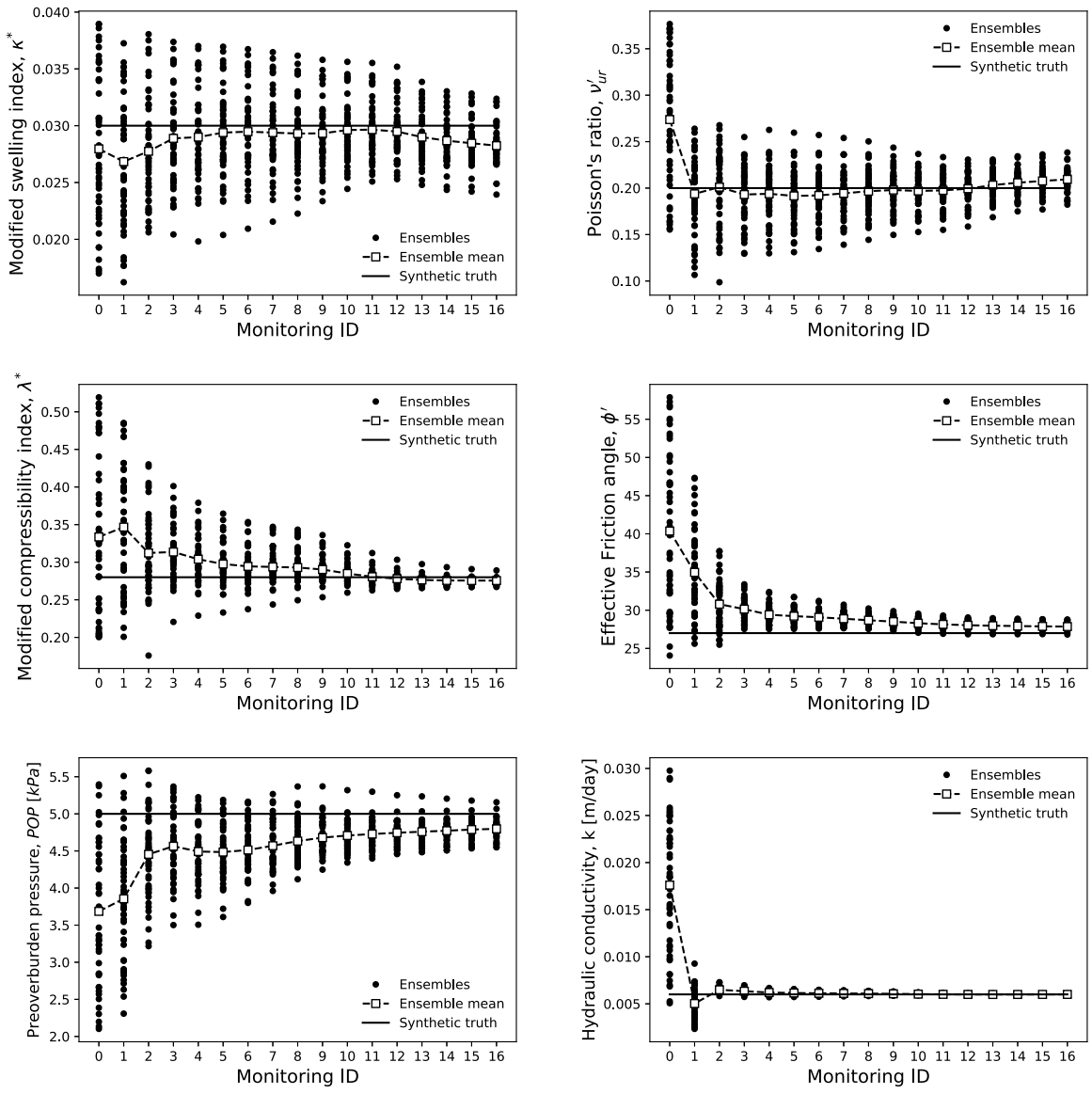


Fig. 5. Convergence of model parameters for the soil using the Ensemble Kalman Filter.

allowing for execution of all possible combinations. The primary benefit of a factorial design is its ability to provide comprehensive data corresponding to various responses observed in each trial, representing the *main effects* of the parameters (Box et al., 2009). In a two-level factorial design involving k input parameters, the main effects are measured by using Eq. (13). In this equation, $\mathfrak{C}$ represents contrast matrix, $\mathfrak{R}$ response vector, $\mathfrak{E}$ vector of effects, and n is the total number of simulations (*e.g.*, Tahershamsi and Dijkstra, 2022).

$$\mathfrak{E} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ \vdots \\ E_{n-1} \\ E_n \end{bmatrix} = \frac{2}{n}\mathfrak{C}^T_{(n\times k)} \cdot \mathfrak{R} = \frac{2}{n}\begin{bmatrix} -1 & -1 & \dots & -1 \\ +1 & -1 & \dots & -1 \\ -1 & +1 & \dots & -1 \\ +1 & +1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & +1 & \dots & +1 \\ +1 & +1 & \dots & +1 \end{bmatrix}^T \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} \tag{13}$$

The levels for this study are set to maintain a relative range of 35% consistently for all parameters, as illustrated in Table 3. All the spatial monitoring points (*p1* to *p8*) mentioned in Section 4.2 are considered for the sensitivity analysis. With regards to the temporal domain, all time periods from Table 2 are selected.

Table 3
Chosen levels of each factor.

Factor	Unit	−	+
λ^*	–	0.224	0.325
κ^*	–	0.020	0.029
ν'_{ur}	–	0.200	0.290
ϕ'	degrees	24.00	35.0
POP	kPa	5.000	7.250
k	m/day	8e−3	11.6e−3

5. Results

5.1. Parameter estimation

Fig. 5 shows the convergence of parameters of the Soft Soil model. The final assimilated ensembles represented by their mean and COV values are summarised in Table 4, while Fig. 6 shows the uncertainty ratio, Z (from Eq. (6)), of each parameter.

The hydraulic conductivity k exhibits the highest rate of convergence, followed by the effective friction angle ϕ', the modified compressibility index λ^* and preoverburden pressure POP. The difference

Table 4
Post-assimilated values of the model parameters.

Parameter	Unit	Synthetic truth	Prior mean	Prior COV	Assimilated mean	Assimilated COV
λ^*	–	0.280	0.333	0.250	0.275	0.013
κ^*	–	0.0300	0.0279	0.250	0.0282	0.061
ν'_{ur}	–	0.200	0.273	0.250	0.209	0.054
ϕ'	degrees	27.00	40.360	0.250	27.857	0.013
POP	kPa	5.000	3.685	0.250	4.797	0.024
k	m/day	0.006	0.0176	0.400	0.0059	0.004

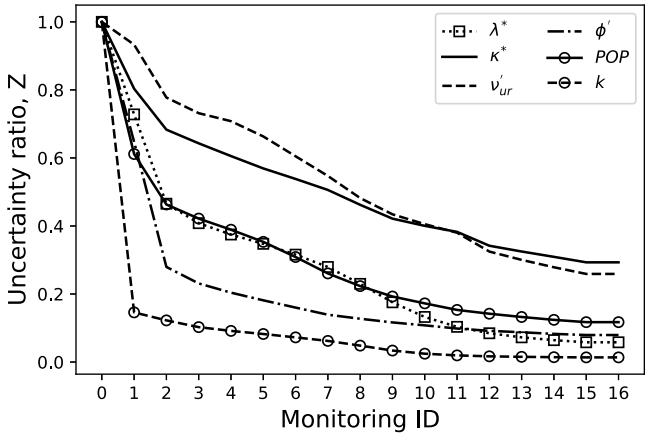


Fig. 6. Uncertainty ratio of model parameters.

in the convergence can be attributed to the varying sensitivity of each parameter to the measurement data included in the Data Assimilation process. Also the change in effective stress level affects the sensitivity of each parameter *e.g.*, the modified swelling index κ^* is only relevant when the effective stress level is below the preconsolidation pressure and that for modified compressibility index λ^* afterwards.

5.2. State estimation

After the first stage of embankment construction, the displacements are monitored and this information is used in the Data Assimilation process to estimate the state, *i.e.*, the displacements of the subsoil. Fig. 7 shows the difference in predicted displacements between the ensemble mean of the prior knowledge of the parameters and those updated with Data Assimilation, for a point 2.0 m below the centre (settlements) and toe (horizontal displacements) of the embankment. The time series shows the mean of the ensemble calculated up to that time when observation is available and then restarted from the initial time period to update the states (as mentioned in Section 3). Using prior knowledge alone, the settlements are overestimated (Fig. 7a) for the complete time period and the horizontal displacements are underestimated after 120 days (Fig. 7b). Using Data Assimilation, the model is able to capture the synthetic true values of the settlement and the horizontal displacement quite accurately. Regarding precision, represented by the standard deviation, the estimate from the settlement of the monitoring points, *p1* to *p4*, ranges from 0.67 to 1.39 mm, while the estimate from the horizontal displacement of points *p5* to *p8* ranges from 0.34 to 0.96 mm showing a high level of precision achieved.

As mentioned in Section 4, the embankment is raised to 3 m after 425 days where the subsequent long-term consolidation is not monitored. The difference in prediction, after the second stage of embankment construction, between the calibrated model and using only prior knowledge is shown in Fig. 8. The results indicate that the model with assimilated parameters captures the response of the embankment in the prediction window more accurately.

Although in this study, the porewater pressure is not included as part of the observation space in the Data Assimilation procedure, the estimation of excess porewater pressure is sufficiently close to the synthetic truth in both the assimilation and prediction window (see Fig. 9). This is reasonable, since a synthetic dataset is created from a model with pre-specified hydraulic and mechanical parameters which, in most cases, is straightforward to back-calculate due to the controlled environment. When dealing with in-situ measurements, however, differences may arise in predictions between mechanical and hydraulic properties due to a lot of hidden uncertainties and various sources of variability which are difficult to capture.

5.3. Effect of different monitoring strategy

As shown in Figs. 5 and 6, there obviously exist differences in the accuracy, precision and rate of convergence between the parameters. Although some parameters fall behind in terms of either accuracy or precision, the state estimate, as shown in Section 5.2 for the vertical and horizontal displacements, is still captured reasonably well. This is due to the fact that the sensitivity of some parameters dictate the accuracy and precision of the state prediction. In order to understand the effect of parameter sensitivity on the performance of Data Assimilation (DA), it is necessary to present the difference in influence between the quantity and type of information. For this, three cases with different monitoring strategies are chosen:

- ***Case-1***: 4 sensors below the centre of the embankment at depths 0.5 m, 1.0 m, 2.0 m and 3.0 m below the ground surface to monitor only the vertical displacement → U_y (monitoring points *p1* to *p4* from Section 4.2).

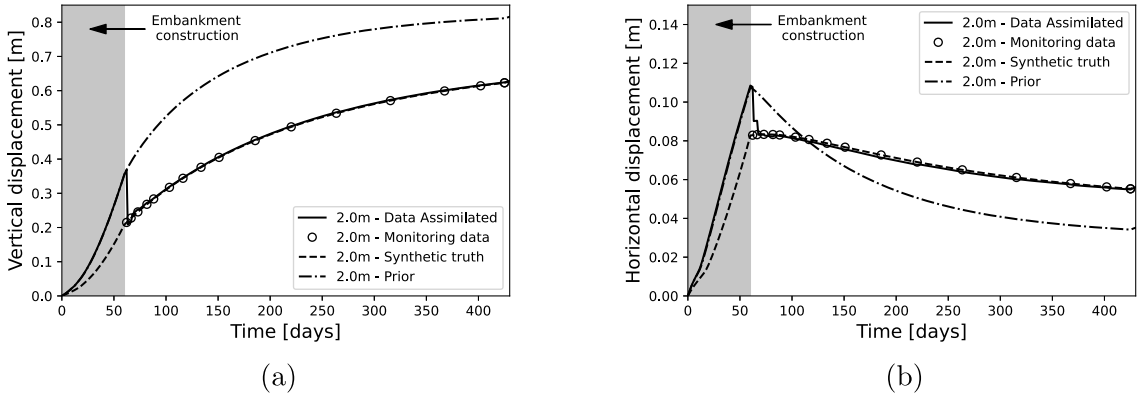


Fig. 7. Estimation of displacements using Data Assimilation in the assimilation window until 425 days at 2 m depth (a) for vertical displacement under the centre of the embankment and (b) for horizontal displacement under the toe of the embankment.

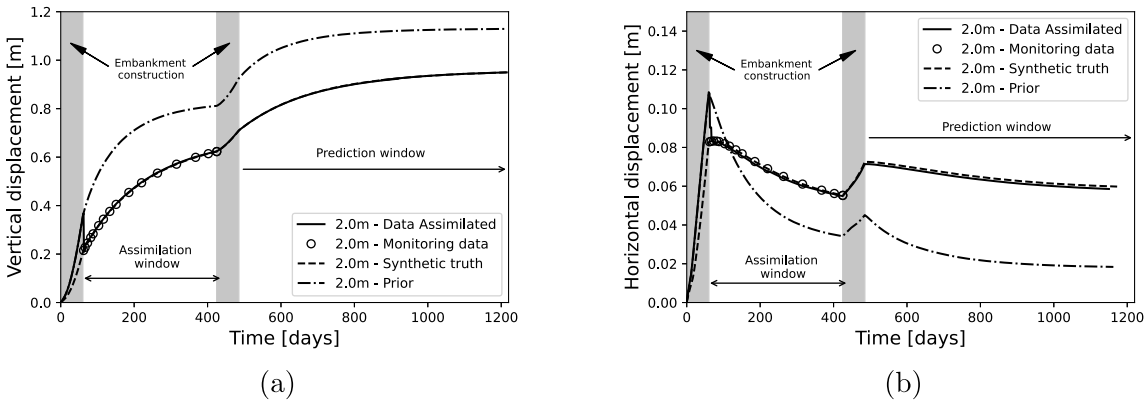


Fig. 8. Estimation of displacements using Data Assimilation in the assimilation and prediction window at 2 m depth (a) for vertical displacement under the centre of the embankment and (b) for horizontal displacement under the toe of the embankment.

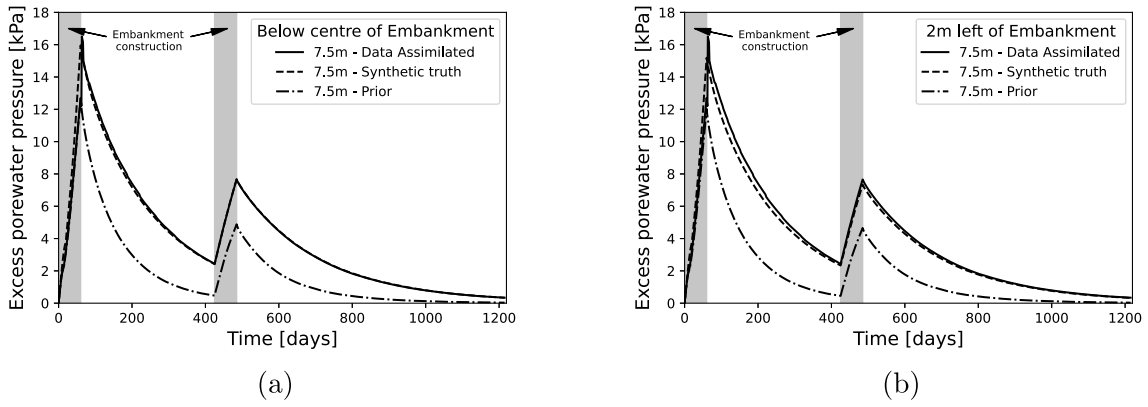


Fig. 9. Estimation of excess porewater pressure at 7.5 m depth (a) under the centre of the embankment and (b) 2 m left of the embankment.

- ***Case-2***: Along with the setup mentioned in Case-1, 4 additional sensors 2.0 m left of the centre of the embankment with the same depth are chosen *i.e.*, all 8 sensors monitor the vertical displacements only → $U_y + U_y$.
- ***Case-3***: The original set-up with 8 sensors as mentioned in Section 4.2 → $U_y + U_x$ (monitoring points *p1* to *p8*).

For all 3 cases, the ensembles generated in Section 4.3 are chosen for the Data Assimilation procedure. In terms of parameter estimation, the final assimilated parameter values from the assimilation window (*i.e.*, up to 425 days) are shown in Table 5. Furthermore, the rate of convergence of all parameters in each case is represented as the uncertainty ratio Z_A (from Eq. (6)) and shown in Fig. 10. A improvement is obtained in terms of precision and accuracy for the model parameters, by increasing the number of sensors (Case-1 → Case-2) but it is not substantial.

In terms of state estimation, Fig. 11 shows the prediction in terms of ensemble mean for Case-2. It can be observed that the vertical displacement and excess porewater pressure are captured reasonably well, while the synthetic truth of the horizontal displacement is not captured, as data on the horizontal displacements is not included in the observation space of the DA algorithm.

The Case-3 monitoring strategy, which is the reference already discussed in Section 4.2, shows significant improvement in terms of both accuracy and precision compared to Cases 1 and 2. Due to the combination of different types of observation, all the parameters achieve better convergence *e.g.*, the effective friction angle and Poisson's ratio are better estimated using Case-3, when included in the observation space in the Data Assimilation procedure. This explains the overall improved convergence for Case-3 compared to the other cases in Fig. 10. The

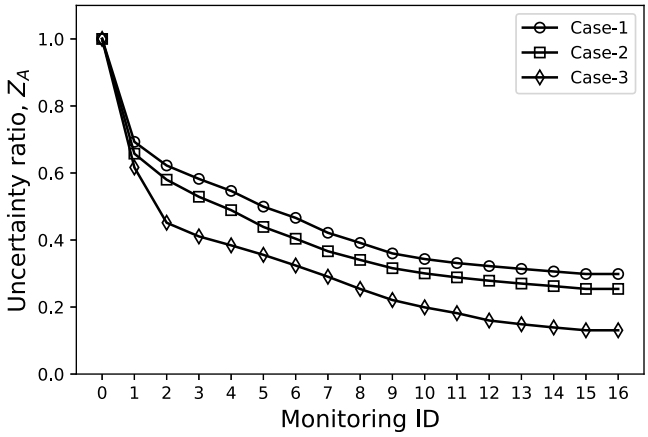


Fig. 10. Uncertainty ratio of model parameters.

behaviour at the toe of the embankment is better captured in the prediction window (also shown in Fig. 8b), due to the inclusion of the horizontal displacement as part of the monitoring scheme.

In this study, the factorial analysis (Section 4.4 for details) is used to evaluate the sensitivity of the parameters. Physical monitoring points *p1* to *p4* are used for analysing the sensitivity of the parameters to the vertical displacement while points *p5* to *p8* are used for the same purpose but for horizontal displacement. Fig. 12 shows the result of the sensitivity of the parameters with respect to vertical displacement U_y for the points below the centre of the embankment (*p1* to *p4*). In

Table 5
Post-assimilated values of the parameters for all cases.

Parameter	Unit	Truth	Prior		Case-1		Case-2		Case-3	
			mean	COV	mean	COV	mean	COV	mean	COV
λ^*	–	0.280	0.333	0.25	0.273	0.020	0.276	0.014	0.275	0.013
κ^*	–	0.030	0.0279	0.25	0.028	0.148	0.0283	0.119	0.0282	0.061
ν'_{ur}	–	0.200	0.273	0.25	0.184	0.203	0.181	0.184	0.209	0.054
ϕ'	degrees	27.00	40.360	0.25	33.582	0.096	33.771	0.075	27.857	0.013
POP	kPa	5.000	3.685	0.25	4.394	0.065	4.448	0.048	4.797	0.024
k	m/day	0.006	0.0176	0.40	0.0062	0.029	0.0063	0.022	0.0059	0.004

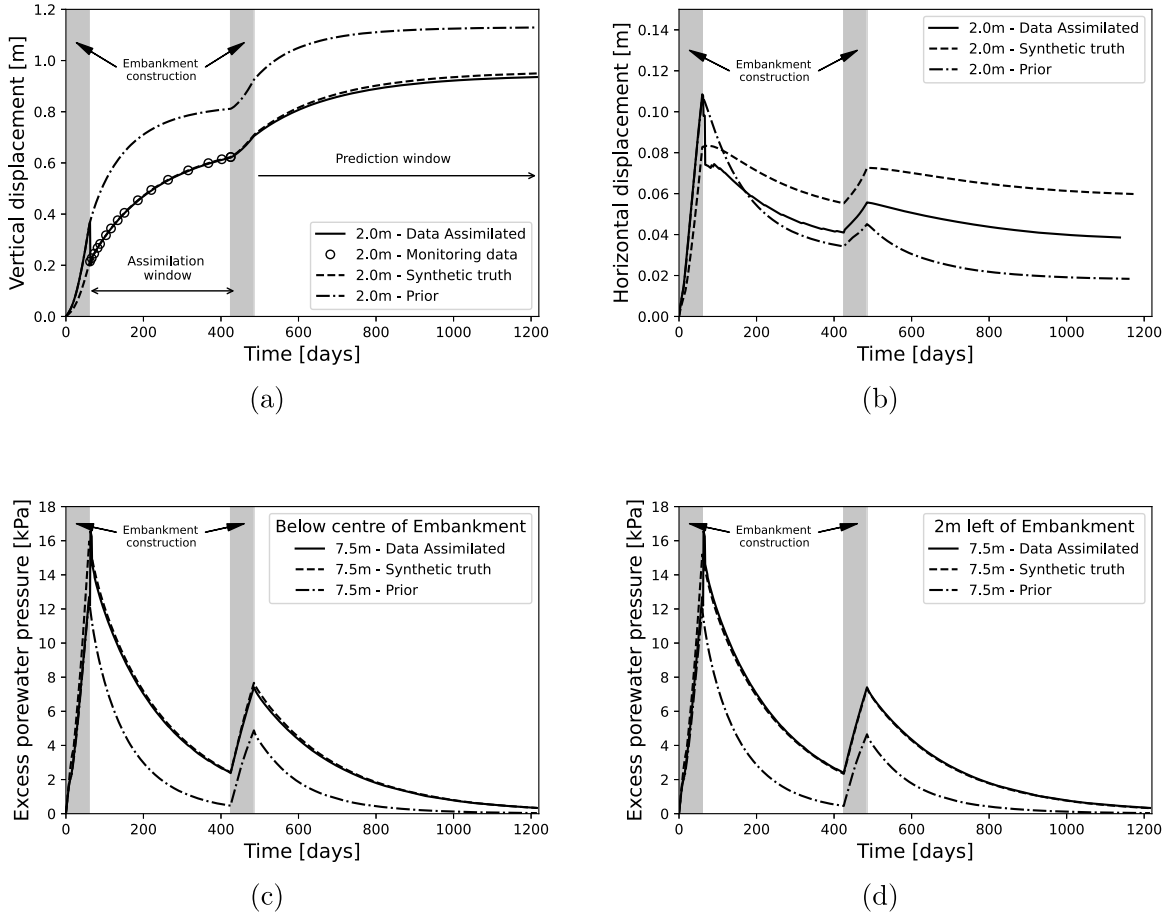


Fig. 11. State estimation of Case-2 at 2.0 m depth for (a) vertical displacement under centre of embankment (b) horizontal displacement at toe of embankment and at 7.5 m depth for excess porewater pressure (c) below the centre of embankment and (d) 2 m left of the embankment.

the initial phase, the hydraulic conductivity k has the largest influence on the vertical displacement. This corroborates well with the rapid convergence of its ensembles when employing Data Assimilation in the initial phase of the simulation (see Fig. 5). However, as shown in Fig. 12, its influence reduces with time and the modified compressibility index λ^* along with the preoverburden pressure (POP) gradually gain high sensitivity on the vertical displacement as time progresses. Hierarchically, this is followed by the effective friction angle ϕ', modified swelling index κ^* and the Poisson's ratio ν'_{ur}. These results are more or less consistent with depth for all physical monitoring points (*p1* to *p4*). Fig. 13 shows the uncertainty ratio for each parameter for Case-1. Clearly, the information of the rate of convergence of each parameter from Data Assimilation correlates well with the hierarchical sensitivity information of the parameters. This shows that the performance of the Data Assimilation procedure is strongly influenced by the spatiotemporal sensitivity of the parameters.

Since the prior knowledge for the parameters in this study are purposely chosen to be far away from the synthetic truth, the accuracy and precision of the relatively low sensitive parameters is heavily affected *e.g.*, the effective friction angle and the Poisson's ratio for Cases 1 and 2, as shown in Table 5, show poor performance. Their final assimilated mean in both cases fail to reach the synthetic true value and the COV is still large after the assimilation process, showing that an increased quantity of sensors is not a complete solution to this problem since the included observation data does not have the necessary information pertaining to those parameters to achieve convergence.

Before analysing the effect of the type of observation, the influence of the choice of the prior distribution of parameters on the results (*i.e.*, Table 5) is investigated. For this, Case-2 is repeated with a well-informed prior distribution for the parameters. A log-normal distribution with mean values closer to the synthetic true value than previously is considered. A COV value of 0.10 consistent for all parameters is chosen for this demonstration (denoted as Prior-2 in Table 6). The previously considered weakly-informed prior distribution is here denoted as Prior-1 for comparison in Table 6. The results show that the final assimilated values for the parameters ϕ' and ν'_{ur} for the

Table 6
Post-assimilated values of the parameters for different initial prior distributions for Case-2.

Parameter	Unit	Truth	Prior-1		Assimilated-1		Prior-2		Assimilated-2	
			mean	COV	mean	COV	mean	COV	mean	COV
λ^*	–	0.280	0.333	0.25	0.276	0.014	0.320	0.10	0.277	0.013
κ^*	–	0.030	0.0280	0.25	0.0283	0.119	0.0280	0.10	0.0250	0.075
ν'_{ur}	–	0.200	0.273	0.25	0.181	0.184	0.230	0.10	0.189	0.057
ϕ'	degrees	27.00	40.36	0.25	33.771	0.075	36.00	0.10	34.174	0.066
POP	kPa	5.000	3.685	0.25	4.448	0.048	4.200	0.10	4.412	0.038
k	m/day	0.006	0.0176	0.40	0.0063	0.022	0.008	0.10	0.0062	0.020

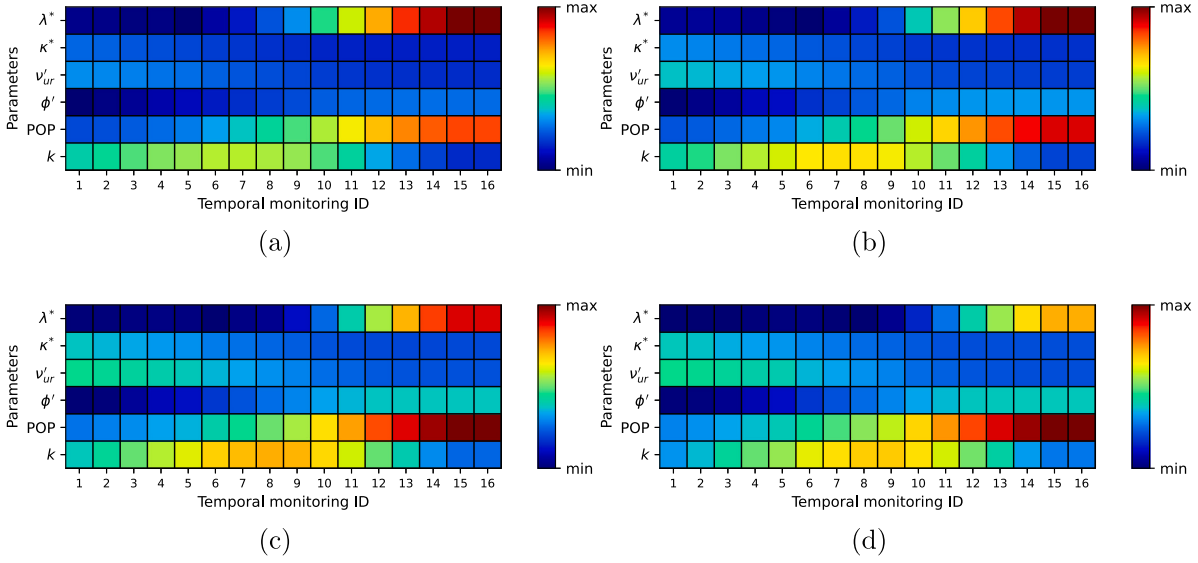


Fig. 12. Sensitivity of parameters with respect to the vertical displacement (U_y) at different time period for physical points below the centre of the embankment (a) p1 [0.5 m] (b) *p2* [1.0 m] (c) *p3* [2.0 m] (d) *p4* [3.0 m].

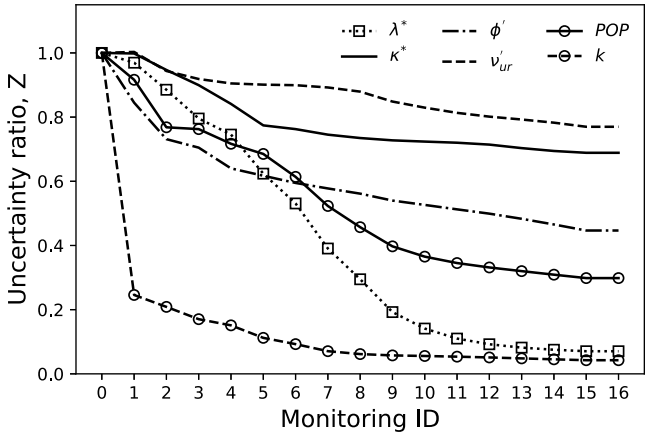


Fig. 13. Uncertainty ratio of model parameters for Case-1.

well-informed prior distribution still fails to converge toward the synthetic truth and there is little to no significant improvement from the previously considered weakly-informed prior distribution.

Although the quantity of sensors between Cases 2 and 3 are the same, the difference in the DA performance is highly apparent and the reason is better explained with the sensitivity results from Figs. 14 and 15. Fig. 14 shows the sensitivity of the parameters with respect to the horizontal displacement (U_x) for multiple points as function of depth below the toe of the embankment (*p5* to *p8*). Contrary to the previous case, parameters k, λ^* and POP show relatively less sensitivity to the horizontal displacement. The Poisson's ratio shows the highest sensitivity for all spatial points considered. The sensitivity of the effective friction angle (ϕ') gradually increases with time but is not consistent with depth for this particular problem. The high sensitivity of ϕ' to the horizontal nodal displacement is also reported by Mohsan et al. (2021) for a slope analysis problem involving different constitutive models. Since the monitoring points *p5* to *p8* are considered in Case-3 as part of its observation space, the higher sensitivity of ϕ' and ν'_{ur} is captured, thereby achieving better convergence than the other two cases, as shown in Fig. 5. Although in Fig. 6, the convergence of ν'_{ur} for Case-3 is shown to be marginal compared to other parameters, this can be attributed to the combined effect between vertical and horizontal displacement. In a relative sense, a significant improvement is achieved, as is evident in Fig. 15 and Table 5.

As mentioned earlier, the sensitivity of the effective friction angle ϕ' is not consistent *i.e.*, its influence on the horizontal displacement decreases with depth as shown in Fig. 14. In view of this, the Case-3 monitoring strategy is modified to study the effect of changing sensor location on the performance of DA, herein denoted as Case-3b. For this, the spatial monitoring points *p1* to *p4* are retained to obtain information on the vertical displacements with depth. However, to monitor the horizontal displacements, the spatial monitoring points are chosen at depths 3 m, 4 m, 5 m and 6 m below the toe of the embankment. As shown in Fig. 15, there is an apparent difference in the rate of convergence, between Case-3 and Case-3b, for parameters ν'_{ur} and ϕ' with the latter showing the expected relative decrease in the rate of convergence.

6. Conclusions

The performance of the Ensemble Kalman Filter, a Data Assimilation algorithm (DA), is analysed in a synthetic experiment. A two-stage embankment construction is simulated with a hydro-mechanically coupled

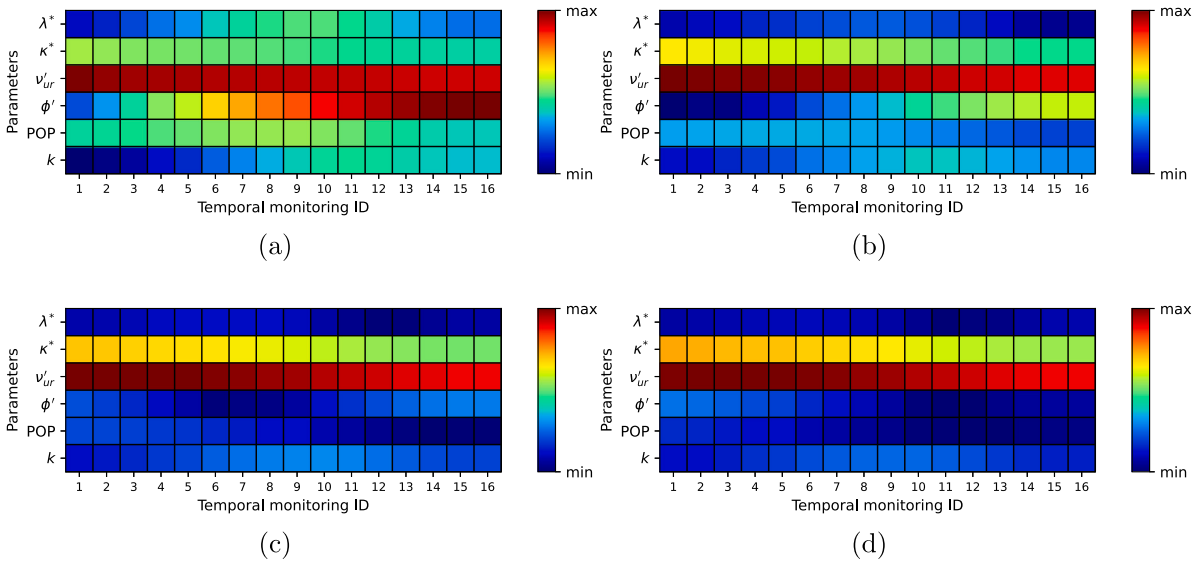


Fig. 14. Sensitivity of model parameters with respect to the horizontal displacement (U_x) at different time period for spatial points below the toe of the embankment (a) *p5* [0.5 m] (b) *p6* [1.0 m] (c) *p7* [2.0 m] (d) *p8* [3.0 m].

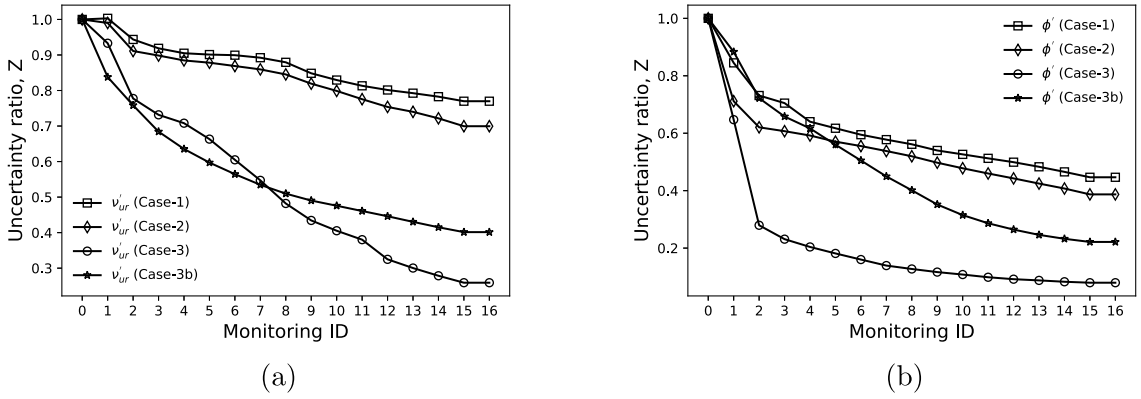


Fig. 15. Uncertainty ratio for the 4 cases for parameters (a) Poisson's ratio (ν'_{ur}) & (b) Effective friction angle (ϕ').

Finite Element (FE) model. The FE model was used to generate the synthetic monitoring data for vertical and horizontal displacements and as the forward model in the DA algorithm. The Soft Soil model has been used to estimate the constitutive behaviour of the subsoil.

The results of the assimilated parameters using Ensemble Kalman Filter exhibit different rate of convergence toward their synthetic true value. For example, in this chosen scenario, the hydraulic conductivity exhibits the highest rate of convergence, followed by the effective friction angle, the modified compressibility index and the preoverburden pressure. With regards to state estimation, the model with the assimilated parameters is able to capture the synthetic true values of the vertical and horizontal displacements with sufficient accuracy and precision in the assimilation window and in the prediction window where monitoring data is not available. In order to investigate the difference in influence between the quantity and type of measurement, different monitoring strategies were chosen in this study. The results indicate that the effective friction angle and Poisson's ratio parameters are better estimated when the horizontal displacement is included along with the vertical displacement in the observation space of the Data Assimilation procedure than for cases hat include only vertical displacement, regardless of the increased quantity of sensors, showing that the latter is not a complete solution to this problem since the included observation data does not have the necessary information pertaining to those parameters to achieve convergence. This finding corroborates well with the results of the global sensitivity analysis performed in this study. The hierarchical sensitivity of the parameters holds a strong influence on their corresponding rate of convergence in the Data Assimilation procedure. The influence on the choice of the prior distribution of the parameters on the Data Assimilation performance has been studied. It is found that there is little to no significant improvement when choosing between log-normal or uniform distribution. The sensitivity of the effective friction angle to the horizontal displacement gradually increases with time but is not consistent with depth in this particular case. In this regard, by modifying the monitoring location for the horizontal displacement, a significant difference in the rate of convergence of the effective friction angle parameter is observed showing the effect of sensor location on the Data Assimilation performance.

It is important to acknowledge that this study focuses solely on a synthetic scenario and makes assumptions on the measurement errors. In future work, the current study needs to be extended to explore a configuration that incorporates more realistic soil properties, such as the spatial variability of the ground conditions, as this can significantly affect the required quantity, quality, type and location of monitoring points.

CRediT authorship contribution statement

Amardeep Amavasai: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization.

Hossein Tahershamsi: Software, Validation, Formal analysis, Writing – original draft. **Tara Wood:** Resources, Supervision, Funding acquisition. **Jelke Dijkstra:** Resources, Supervision, Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A basic version of the python script used in this paper is made available in Github as 542 open source in the link: https://github.com/amaran1988/DA-PLAXIS2D.git.

Acknowledgements

The financial support from FORMAS under grant no. 2020/00220 and Nordforsk, Norway (project #98335 NordicLink) is greatly acknowledged. The work is performed as part of Digital Twin Cities Centre that is supported by Sweden's Innovation Agency VINNOVA.

References

Bocquet, M., Sakov, P., 2013. Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlinear Processes Geophys. 20, 803–818.
Box, G.E.P., Hunter, J.S., Hunter, W.G., 2009. Statistics for Experimenters; Design, Innovation, and Discovery. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 1–655.
Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. On the analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126.
Carrassi, A., Bocquet, M., Bertino, L., Evensen, G., 2017. Data assimilation in the geosciences - An overview on methods, issues and perspectives. Wiley Interdiscip. Rev. Clim. Change 9.
Chatzi, E.N., Smyth, A.W., 2009. The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing†. Struct. Control Health Monit. 16, 99–123.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162.
Iglesias, M.A., Law, K.J.H., Stuart, A.M., 2013. Ensemble Kalman methods for inverse problems. Inverse Problems 29 (4), 045001.
Juang, C., Luo, Z., Atamturktur, S., Huang, H., 2013. Bayesian updating of soil parameters for braced excavations using field observations. J. Geotech. Geoenviron. Eng. 139.
Kálmán, R.E., 1960. A new approach to linear filtering and prediction problems. ASME Journal of Basic Engineering 82, 35–45.
Karstunen, M., Amavasai, A., 2017. BEST SOIL : Soft Soil Modelling and Parameter Determination. Technical Report, Chalmers University of Technology.
Liu, Z., Choi, J.C., Lacasse, S., Nadim, F., 2018. Uncertainty analyses of time-dependent behaviour of Ballina test embankment. Comput. Geotech. 93, 133–149, Ballina Embankment Prediction Symposium.
Mavritsakis, A., 2017. Evaluation of Inverse Analysis Methods with Numerical Simulation of Slope Excavation (Msc thesis), Delft University of Technology.
Mohsan, M., Vardon, P., Vossepoel, F., 2021. On the use of different constitutive models in data assimilation for slope stability. Comput. Geotech. 138, 104332.
Plaxis, 2015. PLAXIS material models manual. Plaxis.
Roscoe, K.H., Burland, J.B., 1968. On the generalized stress-strain behaviour of wet clay. In: Heyman, J., Leckie, F. (Eds.), Engineering Plasticity. Cambridge University Press, pp. 553–609.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons, Ltd, Chichester, UK, p. 312.
Schillings, C., Stuart, A., 2016. Analysis of the ensemble Kalman filter for inverse problems. SIAM J. Numer. Anal. 55.
Tahershamsi, H., Dijkstra, J., 2022. Using experimental design to assess rate-dependent numerical models. Soils Found. 62 (6), 101244.
Tao, Y., Sun, H., Cai, Y., 2020. Predicting soil settlement with quantified uncertainties by using ensemble Kalman filtering. Eng. Geol. 276, 105753.
Tao, Y., Sun, H., Cai, Y.-q., 2021. Bayesian inference of spatially varying parameters in soil constitutive models by using deformation observation data. Int. J. Numer. Anal. Methods Geomech. 45.
Tao, Y., Yu, M., Sun, H., 2022. A comparison between EnKF and MCMC-based Bayesian updating for consolidation settlement prediction. In: International Symposium for Geotechnical Safety & Risk. pp. 897–902.
Trudinger, C., Raupach, M., Rayner, P., Enting, I., 2008. Using the Kalman filter for parameter estimation in biogeochemical models. Environmetrics 19, 849–870.
Van Rossum, G., Drake Jr., F.L., 1995. Python Reference Manual. Centrum voor Wiskunde en Informatica Amsterdam.
Vardon, P.J., Liu, K., Hicks, M.A., 2016. Reduction of slope stability uncertainty based on hydraulic measurement via inverse analysis. Georisk: Assess. Manag. Risk Eng. Syst. Geohazards 10 (3), 223–240.
Verlaan, M., Heemink, A., 2001. Nonlinearity in data assimilation applications: A practical method for analysis. Mon. Weather Rev. 129.
Wu, T.H., Zhou, S.Z., Gale, S.M., 2007. Embankment on sludge: predicted and observed performances. Can. Geotech. J. 44 (5), 545–563.
Zhang, J., Zhang, L., Tang, W., 2010. Back analysis of slope failure with Markov chain Monte Carlo simulation. Comput. Geotech. 37, 905–912.