Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_17-009-cd

ISGSR 2022

A Comparison between EnKF and MCMC-Based Bayesian Updating for Consolidation Settlement Prediction

Yuanqin Tao[1], Mengfei Yu[2], and Honglei Sun[3]

[1]Institute of Geotechnical Engineering, College of Civil Engineering, Zhejiang University of Technology,
288 Liuhe Road, Hangzhou, P.R. China.
E-mail: taoyuanqin@zju.edu.cn
[2]Institute of Geotechnical Engineering, College of Civil Engineering, Zhejiang University of Technology,
288 Liuhe Road, Hangzhou, P.R. China.
E-mail: yumengfei@zjut.edu.cn
[3]Institute of Geotechnical Engineering, College of Civil Engineering, Zhejiang University of Technology,
288Liuhe Road, Hangzhou, P.R. China.
E-mail: sunhonglei@zju.edu.cn

Abstract: Consolidation settlement is usually predicted based on the geotechnical parameters obtained by the field and laboratory tests. However, such predictions usually deviate from field monitoring data due to uncertainties in parameter selection. Bayesian methods provide an effective way to update the geotechnical parameters and improve the prediction accuracy by incorporating the monitoring data. Markov chain Monte Carlo (MCMC) is a popular method to derive the posterior distribution of soil parameters. It can achieve a rigorous sampling but generally requires tens of thousands of forward model calculations. Ensemble Kalman filter (EnKF), as an alternative, can efficiently deal with recursive updating based on the sequential monitoring data, but accompanied by some limitations due to its Gaussian assumptions and linear update. This paper evaluates the performance of EnKF and MCMC methods for identifying the parameters and updating the predictions of consolidation settlement througha laboratory test. The results show that the EnKF and MCMC can result in a consistent estimation of the mean values of the updated soil parameters and settlements, while the uncertainties obtained from EnKF are overestimated.

Keywords: Inverse analysis; ensemble Kalman filter; Bayesian updating; consolidation settlement.

1 Introduction

The performance of geotechnical models relies significantly on the values of soil parameters. However, selecting the values of soil parameters involves strong subjectivity from the predictors, resulting in a significant degree of uncertainty(Doherty and Bransby 2021). Bayesian methods provide aneffective way to make use of observation data to release such subjectivity during the parameter selection.

The Bayesian updating approach conjugated with Markov chain Monte Carlo (MCMC)sampling has been widely used in geotechnical problems (Kelly and Huang 2015; Wang and Cao 2013; Zhang et al. 2010). MCMC can provide a rigorous sampling and is regarded as the gold standard of the Bayesian method. However, it requires tens of thousands of forward model calculations involved in the likelihood. Recently, the ensemble Kalman filter (EnKF) has gained attentionin updatingthe soil parameters (Ju et al. 2020; Tao et al. 2020, 2021; Vardon et al. 2016). EnKF is an invariant of the Kalman filter for nonlinear problems. It is computationally efficient as the required number of samples is significantly smaller compared to that of MCMC. However, the deviation of EnKF assumes that the prior distribution and the observation error follow the Gaussian distribution. Moreover, the parameters are updated based on a linear "shift" between the observation and the prior estimation (Katzfuss et al. 2016), using a weight represented by the Kalman matrix. As a result, EnKF only samples from the true posterior distribution in linear problemsbut provides an approximate solution for nonlinear problems. No work has been conducted to assess the impacts of these assumptions on the posterior estimations of soil parameters and geotechnical responses.

The objective of this study is to compare the performance of EnKF and MCMC from several perspectives, including theaccuracy of the prediction mean, the uncertainty quantification, and the computational cost. The paper starts with an introduction of sequential updating, followed by the descriptions of the augmented EnKF and restart EnKF. The EnKF and MCMC-based sequential updating methods are illustrated through a laboratory consolidation test under vacuum preloading combined with vertical drain.

2 Methodology

2.1Bayesian updating framework

Consolidation settlement is a time-dependent processwhere the observation data is obtained sequentially. There are two strategies to incorporate the available observation data in Bayesian updating, i.e., using all the data as a whole and usingeach piece of data one by one(Li et al. 2016).The formerrequires a higher computational cost because the model responses need to be calculated and saved at every observation time. It may also result in numerical problems when the number of observations is large so that the product of likelihood becomes too small. Thus, this study uses the observation data one by one, which is called sequential updating. The posterior distribution of soil parameters **x** is written as

$$p(\mathbf{x} \mid \mathbf{d}_{\mathrm{obs},k}) \propto p(\mathbf{d}_{\mathrm{obs},k} \mid \mathbf{x}) p(\mathbf{x} \mid \mathbf{d}_{\mathrm{obs},k\text{-}1}) \tag{1}$$

where the posterior distribution $p(\mathrm{x} \mid \mathrm{d}_{\mathrm{obs},k\text{-}1})$ obtained from the previous update step $k-1$ is used as the prior distribution in the current step k. In geotechnical problems, the analytical solution of the posterior distribution is frequently unavailable; instead, the Markov chain Monte Carlo (MCMC) algorithm is adopted to derive the posterior distribution.

2.2Sequential updating based on ensemble Kalman filter

Ensemble Kalman filter(EnKF) is an approximate Bayesian method (Xiao et al. 2016) forsequential updating.Different from MCMC which is applicable in any complex posterior distributions, the deviation of EnKF assumes that the prior distribution and observation error follow Gaussian distributions. EnKF is initially proposed to update the system state (e.g., settlement s)and then extended to the parameter-state estimation (e.g., soil parameters θ and settlement s). In the parameter-state estimation problems, the original state is usually defined to include the parameters, leading to an augmented state x = [s, θ]. The EnKF method begins with an initial sampling step. The initial state vectors are generated based on the prior distribution and form the initial ensemble. After the initial sampling step, alternate prediction steps and analysis steps are carried out as follows:

(a) Prediction step. The state variable and the parameters updated from the previous analysis step $t-1$or generated from the initial sampling step are used to predict the state of the system at the current time step t. The superscripts “f” and “a” represent forecast and analysis, respectively.

$$s_t^{\mathrm{f}} = f\left(\mathbf{x}_{t-1}^{\mathrm{a}}\right) = f\left(s_{t-1}^{\mathrm{a}}, \boldsymbol{\theta}_{t-1}^{\mathrm{a}}\right) \tag{2}$$

The parameters are also “predicted” at time t by taking the same value as the updated parameters at $t-1$.

$$\boldsymbol{\theta}_t^{\mathrm{f}} = \boldsymbol{\theta}_{t-1}^{\mathrm{a}} \tag{3}$$

The predicted values of the state variable and parameters constitute the forecasted state vector $\mathbf{x}_t^{\mathrm{f}}$.

$$\mathbf{x}_t^{\mathrm{f}} = \left(s_t^{\mathrm{f}}, \boldsymbol{\theta}_t^{\mathrm{f}}\right) \tag{4}$$

(b) Analysis step. If observation data is available at time t, the forecasted state vector is updated based on Eq. (5) and (6):

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}\left(\mathbf{d}_t - \mathbf{H}\mathbf{x}_t^{\mathrm{f}}\right) \tag{5}$$

$$\mathbf{K} = \mathbf{P}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}}\right)^{-1} \tag{6}$$

where d_t denotes the observation data at time t; K is the Kalman matrix that serves as a weight between the prior information and the observation; R is the matrix of observation error; H is an observation matrix that transforms the augmented state vector to the observation space. It can be seen from Eq. (5) that EnKF performs a linear “shift” to update the state variable and the parameters. The reason for state augmentation is to avoid a rerun ofthe forward model from time 0. It is beneficial to computational efficiency but may result in statistical inconsistency in nonlinear problems (Thulin et al. 2007). In this study, considering the forward model is a series of explicit equations and is fast to compute, the restart EnKF is used rather than the standard EnKF for the parameter-state estimation problem. The forecast equation in Eq. (2) is changed to the following Eq. (7), where the state at time t is calculated by using the state value at time 0 and the updated parameters from the last analysis step $t-1$. The restart EnKF is used in this study except for specific instructions.

$$s_t^{\mathrm{f}} = f\left(s_0, \boldsymbol{\theta}_{t-1}^{\mathrm{a}}\right) \tag{7}$$

3 Case study

This study uses a consolidation prediction example to illustrate the sequential updating based on MCMC and EnKF. Figure1 shows the consolidation model under vacuum preloading combined with vertical drain.The flow is not allowed to cross the boundaries of until cell, and only the radial (horizontal) flow is considered because of the long vertical drain. As shown in Figure 1, the central gray cylinder represents the equivalent drainage area of the vertical drain, while the middle orange layer is the smear zone,representing the disturbed area after inserting the vertical drain. The outer layer is the undisturbed zone. Parameters d_w, d_s, and d_e represent the diameters of the equivalent drain, smear zone, and influence zone, respectively. The horizontal permeability coefficients in the smear zone and the undisturbed zone are k_s and k_h, respectively.

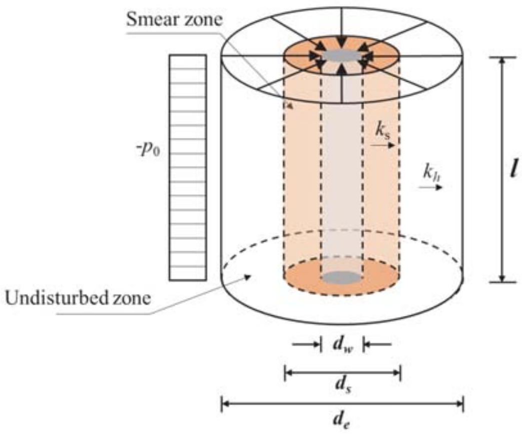


Figure 1. Model geometry

An analytical solution from Indraratna et al. (2005)is adopted to calculate the excess pore-water pressure u as follow:

$$u = \frac{(1+k_1)p_0}{2}\exp(\frac{-8T_h}{\mu}) - \frac{(1+k_1)p_0}{2} \tag{8}$$

wherek_1 is the ratio between vacuum pressure at the top and thebase of the vertical drain,which is simply set to 1 as the well-resistance effect is not considered in this study; $p_0$is the applied vacuum pressure; $T_h$is the dimensionless time factor for horizontal drainage, which is calculated as Eq. (9);μ is a group of parameters representing the geometry of the vertical drain system and smear effect, as shown in Eq. (10).

$$T_h = \frac{c_h t}{d_e^{\,2}} \tag{9}$$

where c_h is the coefficient of consolidation for horizontal drainage, $c_h = \dfrac{k_h}{\gamma_w m_v}$, and $\gamma_w$is the unitweight of water.

$$\mu = \ln(\frac{n}{s'}) + \frac{k_h}{k_s}\ln(s') - \frac{3}{4} \tag{10}$$

where $n = \dfrac{d_e}{d_w}$; and $s' = \dfrac{d_s}{d_w}$. The calculation equation for settlement is finally written as Eq. (11):

$$s = m_v(u_0 - u)l \tag{11}$$

where m_v is the coefficient of volume compressibility,$u_0$is the initialexcess pore-water pressure, and$l$ is the thickness of the layer.

3.1 Computational setup

The sequential updating for consolidation settlement is illustrated through a laboratory test(Sun et al. 2021). All the parameters are set to the same values asSun et al. (2021), which is summarized in Table 1. Among the parameters, three key soil parameters are selected to be updated based on the sequential observation data. They are the horizontal permeability coefficient in the undisturbed zonek_h, the horizontal permeability coefficient in smear zonek_s, and the coefficient of volume compressibility m_v.The prior distributions of three parameters are listed in Table 2. The mean values of parameters are set to the values measured by the laboratory test, while the coefficients of variance (COVs) are set to 0.4.

Table 1. Values of model parameters

Parameter	d_w(m)	d_s (m)	d_e (m)	l (m)	γ_w (kN/m^3)	p_0 (kPa)	u_0 (kPa)	k_h (m/d)	k_s (m/d)	m_v (1/kPa)
Value	0.066	0.174	0.5	0.56	10	85	0	0.0131	4.37×10^{-5}	0.0085

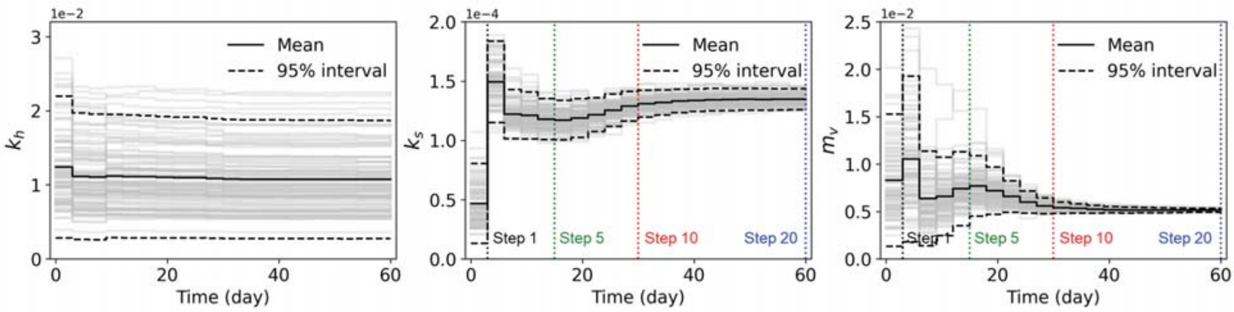
The observation data is the settlement measured every three days. The standard deviation of observation error is assumed as 5% of the measured values. As for the other parameters of sequential updating algorithms, the ensemble size of EnKF is set to 100, and the EnKF is performed ten times with different initial ensembles to test the degree of randomness. In the Metropolis-Hastings method, the total length of the Markov chain is 30000. After the burn-in period, one is taken of every ten samples as a posterior sample.

Table 2. Prior distributions

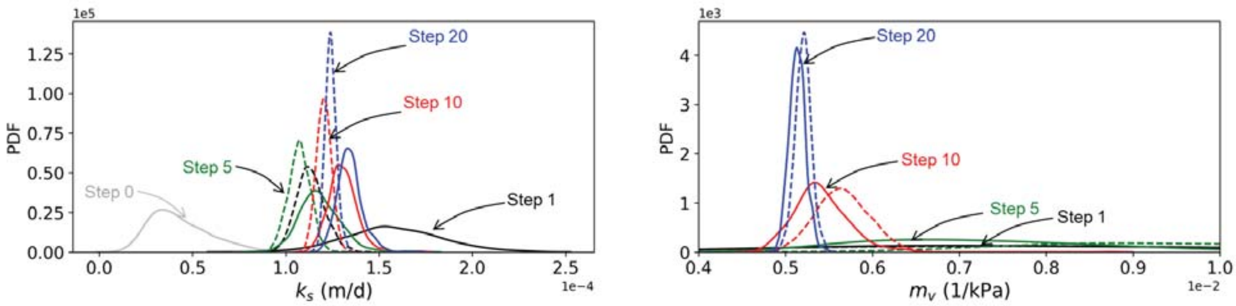
Parameter	Distribution	Mean	COV
k_h (m/d)	Lognormal	0.0131	0.4
k_s (m/d)	Lognormal	4.37×10^{-5}	0.4
m_v(1/kPa)	Lognormal	0.0085	0.4

3.2 Updated parameters

The updating process of parameters is shown in Figure 2. Figure 2(a) shows the convergence history of three parameters using EnKF. The intervals of thehorizontal permeability coefficient in the smear zonek_s and the coefficient of volume compressibility $m_v$shrink quicklyby incorporating observation data, while the horizontal permeability coefficient in the undisturbed zonek_h still has significant uncertainty. Figure 2(b) compares the probability density functions (PDFs) obtained from EnKF and MCMC at assimilation steps 1, 5, 10, and 20. The step number represents the number of assimilated observations. Thus, the PDF at step 0 denotes the prior distribution. The uncertainties of soil parameters are reduced as more observations are incorporated, no matter whether EnKF or MCMC is used. In general, EnKF can provide a similar mean but overestimates the variance compared to MCMC. It should be noted that the PDF of k_s obtained from EnKF considerably deviates from that of MCMC. This is because the single linear update in the assimilation step results in an overcorrection if a significant discrepancy exists between the prediction and the observation. In our case, the prior distribution of soil parameters provides a prediction quite far away from the observation (Figure 3). As a result, the update equation shown by Eq. (7) leads to a large linear shift of the parameters. This kind of overcorrection will be less of a problem if timely dense observations are available.



(a) Convergence history of three parameters using EnKF

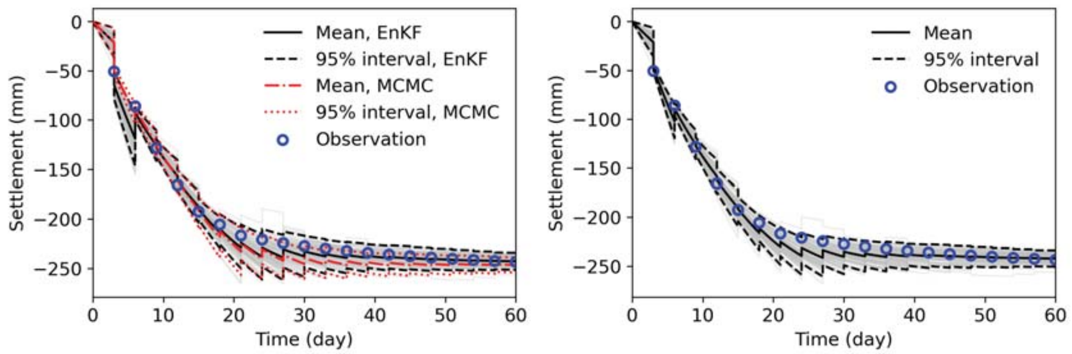


(b) Comparison of PDFs between EnKF and MCMC. The solidline and the dashed line represent the updated PDFs from EnKF and MCMC, respectively.

Figure 2. Updating process of parameters

3.3 Settlement prediction

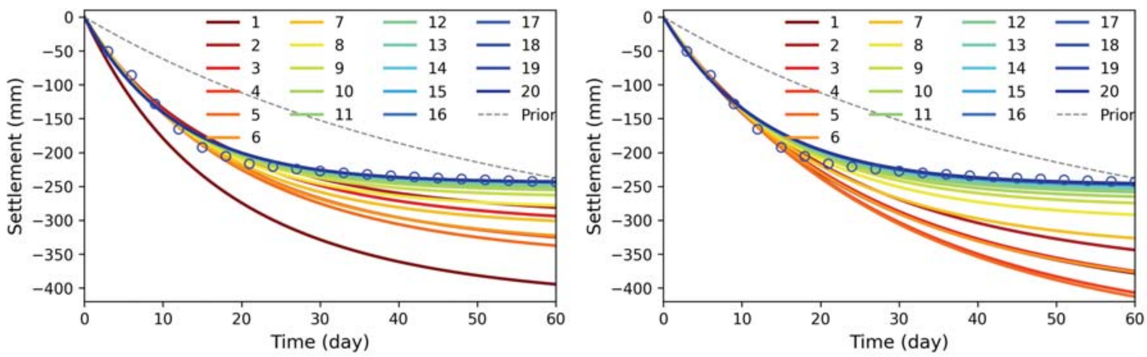
Figure 3 shows the sequential updating of settlement using EnKF and MCMC. Similar to the results of PDFs of soil parameters, a significant overcorrection appears after incorporating the first observation data at day 3if EnKF is used. However, such overcorrection seems to be alleviated in this example when state augmentation is adopted. In general, the prediction mean of settlement is pretty close in both EnKF and MCMC, but the scattering from EnKF is a litter larger than that of MCMC.



(a) Restart EnKF (b) EnKF using augmented state

Figure 3. Sequential updating ofsettlement using EnKF and MCMC

Figure 4 shows the effects of the number of observations on the prediction mean of settlement. The discrepancy between the prediction and the observation data does not decrease monotonically with the increase in the number of observations. In general, the parameters updated from theobservations in the early days overestimate the long-term settlement (e.g., day 60). This is understandable because the material used in the consolidation test is the dredged slurry. Its consolidation is quicker in the early stage due to its initial high permeability and compressibility.



(a) Restart EnKF (b) MCMC

Figure 4. Prediction mean of settlement using different numbers of observations

Figure5shows the settlement prediction at the last observation time (i.e., day 60). The EnKF is performed ten times with different initial ensembles generated from the same prior distribution. It can be seen that the settlement predictions from different initial ensembles are inconsistent. But such fluctuation can be alleviated as more observations stream in.

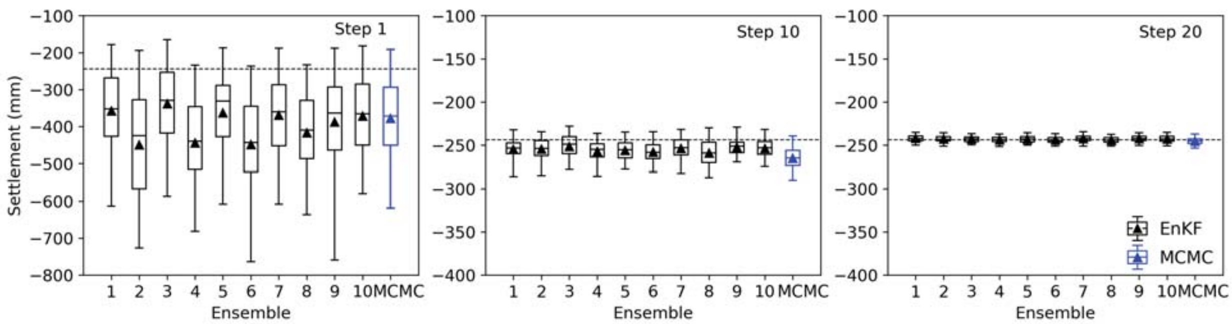


Figure 5. Prediction of settlement onday 60

4 Summary and Conclusions

This study evaluates the ability of EnKF and MCMC for sequential updating of consolidation settlement under vacuum preloading combined with vertical drain. The MCMC-based method is widely used in Bayesian updating for geotechnical problems, while the EnKF has gained attentionrecently due to its high efficiency. The performance of two sequential updating methods is evaluatedusing a laboratory test. Taking the result of MCMC as a benchmark, EnKF provides similar mean values of posterior distributions of soil parameters but overestimates the uncertainty. Moreover, it is possible for EnKF to result in overcorrections of parameters and settlement if observation data is not timely dense, andsuch overcorrection can be alleviated with an increase in the number of observations. In terms of computational efficiency, it takes ⮌ (10^2) and ⮌ (10^5) required samples for EnKF and MCMC to incorporateone piece of observation data, respectively.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 52078464 and U2006225).

References

Doherty, J. P., and Bransby, M. F. (2021). A data-driven approach for predicting the time-dependent settlement of embankments on soft soils. *Géotechnique,* 71(11), 1014-1027.

Indraratna, B., Rujikiatkamjorn, C., and Sathananthan, I. (2005). Analytical and numerical solutions for a single vertical drain including the effects of vacuum preloading. *Canadian Geotechnical Journal,* 42(4), 994-1014.

Ju, L.Y., Miao, C., Cao, Z.J., Hubbard, P., Soga, K., and Li, D. Q. (2020). Uncertainty Quantification of Soil Total Unit Weight Based on Random Field Model and Linear Dynamic System: A Comparative Study.*Geo-Congress 2020, ASCE, Minneapolis, Minnesota*, 558-568.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2016). Understanding the Ensemble Kalman Filter. *The American Statistician,* 70(4), 350-357.

Kelly, R., and Huang, J. (2015). Bayesian updating for one-dimensional consolidation measurements. *Canadian Geotechnical Journal,* 52(9), 1318-1330.

Li, X. Y., Zhang, L. M., and Jiang, S. H. (2016). Updating performance of high rock slopes by combining incremental time-series monitoring data and three-dimensional numerical analysis. *International Journal of Rock Mechanics and Mining Sciences,* 83, 252-261.

Sun, H., Lu, Y., Pan, X., Shi, L., and Cai, Y. (2021). The effect of initial water content on the consolidation of dredged slurry under vacuum preloading. *Rock and Soil Mechanics,* 42(11), 3029-3040.

Tao, Y., Sun, H., and Cai, Y. (2020). Predicting soil settlement with quantified uncertainties by using ensemble Kalman filtering. *Engineering Geology,* 276, 105753.

Tao, Y., Sun, H., and Cai, Y. (2021). Bayesian inference of spatially varying parameters in soil constitutive models by using deformation observation data. *International Journal for Numerical and Analytical Methods in Geomechanics,* 45(11), 1647-1663.

Thulin, K., Li, G., Aanonsen, S., and Reynolds, A. C. (2007). Estimation of initial fluid contacts by assimilation of production data with EnKF. *SPE annual technical conference and exhibition, OnePetro*.

Vardon, P. J., Liu, K., and Hicks, M. A. (2016). Reduction of slope stability uncertainty based on hydraulic measurement via inverse analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards,* 10(3), 223-240.

Wang, Y., and Cao, Z. (2013). Probabilistic characterization of Young's modulus of soil using equivalent samples. *Engineering Geology,* 159, 106-118.

Xiao, H., Wu, J. L., Wang, J. X., Sun, R., and Roy, C. J. (2016). Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier-Stokes simulations: A data-driven, physics-informed Bayesian approach. *Journal of Computational Physics,* 324, 115-136.

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